



University of Kentucky  
UKnowledge

---

Theses and Dissertations--Statistics

Statistics

---

2011

## PARAMETRIC ESTIMATION IN COMPETING RISKS AND MULTI-STATE MODELS

Yushun Lin

University of Kentucky, [yslin8@gmail.com](mailto:yslin8@gmail.com)

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

---

### Recommended Citation

Lin, Yushun, "PARAMETRIC ESTIMATION IN COMPETING RISKS AND MULTI-STATE MODELS" (2011).  
*Theses and Dissertations--Statistics*. 1.  
[https://uknowledge.uky.edu/statistics\\_etds/1](https://uknowledge.uky.edu/statistics_etds/1)

This Doctoral Dissertation is brought to you for free and open access by the Statistics at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Statistics by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Yushun Lin, Student

Dr. Richard J. Kryscio, Major Professor

Dr. Arne Bathke, Director of Graduate Studies

PARAMETRIC ESTIMATION IN COMPETING RISKS AND MULTI-STATE  
MODELS

---

DISSERTATION

---

A dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy in the  
College of Arts and Sciences  
at the University of Kentucky

By  
Yushun Lin  
Lexington, Kentucky

Director: Dr. Richard J. Kryscio, Professor of Statistics  
Lexington, Kentucky

2011

Copyright© Yushun Lin 2011

## ABSTRACT OF DISSERTATION

### PARAMETRIC ESTIMATION IN COMPETING RISKS AND MULTI-STATE MODELS

The typical research of Alzheimer's disease includes a series of cognitive states. Multi-state models are often used to describe the history of disease evolvement. Competing risks models are a sub-category of multi-state models with one starting state and several absorbing states.

Analyses for competing risks data in medical papers frequently assume independent risks and evaluate covariate effects on these events by modeling distinct proportional hazards regression models for each event. Jeong and Fine (2007) proposed a parametric proportional sub-distribution hazard (SH) model for cumulative incidence functions (CIF) without assumptions about the dependence among the risks. We modified their model to assure that the sum of the underlying CIFs never exceeds one, by assuming a proportional SH model for dementia only in the Nun study. To accommodate left censored data, we computed non-parametric MLE of CIF based on Expectation-Maximization algorithm. Our proposed parametric model was applied to the Nun Study to investigate the effect of genetics and education on the occurrence of dementia. After including left censored dementia subjects, the incidence rate of dementia becomes larger than that of death for age  $< 90$ , education becomes significant factor for incidence of dementia and standard errors for estimates are smaller.

Multi-state Markov model is often used to analyze the evolution of cognitive states by assuming time independent transition intensities. We consider both constant and duration time dependent transition intensities in BRAiNS data, leading to a mixture of Markov and semi-Markov processes. The joint probability of observing a sequence of same state until transition in a semi-Markov process was expressed as a product of the overall transition probability and survival probability, which were simultaneously modeled. Such modeling leads to different interpretations in BRAiNS study, i.e., family history, APOE4, and sex by head injury interaction are significant factors for transition intensities in traditional Markov model. While in our semi-Markov model, these factors are significant in predicting the overall transition probabilities, but none of these factors are significant for duration time distribution.

KEYWORDS: cumulative incidence function, interval censored data, semi-Markov, competing risks, multi-state model

Author's Signature: Yushun Lin

Date: December 13, 2011

PARAMETRIC ESTIMATION IN COMPETING RISKS AND MULTI-STATE  
MODELS

By  
Yushun Lin

Director of Dissertation: Dr. Richard J. Kryscio

Director of Graduate Studies: Dr. Arne Bathke

Date: December 13, 2011

## ACKNOWLEDGMENTS

I would like to express my appreciation to my advisor Dr. Richard Kryscio for his advice and support during my graduate studies. His wisdom, guidance and excellent suggestions helped me better understand and finish my dissertation.

I want to thank my committee members, Dr. William Griffith, Dr. Craig Miller, Dr. Arne Bathke, and Dr. Yanbing Zheng for serving on my Supervisory Committee. I am grateful to them for their time, and the careful and critical reading of this dissertation. I also thank Erin L Abner for valuable comments and discussion on the BRAiNS study.

I feel a deep sense of gratitude to my family and friends for their constant love, dedication and support. They are the source of my happiness and motivation behind my achievements.

## TABLE OF CONTENTS

Acknowledgments . . . . .	v
Table of Contents . . . . .	vi
List of Figures . . . . .	vii
List of Tables . . . . .	viii
Chapter 1 Introduction . . . . .	1
1.1 Overview . . . . .	1
1.2 Competing Risks Models . . . . .	2
1.3 Multi-state Models . . . . .	9
1.4 Estimation Methods . . . . .	16
1.5 Outline of the Dissertation . . . . .	25
Chapter 2 Competing Risks Modeling . . . . .	27
2.1 Introduction . . . . .	27
2.2 Model Setup . . . . .	30
2.3 Simulation Studies . . . . .	34
2.4 Application To The Nun Study . . . . .	40
2.5 Discussion . . . . .	47
Chapter 3 Multi-state Modeling . . . . .	50
3.1 Introduction . . . . .	50
3.2 Model Specification . . . . .	54
3.3 Estimation and Model Selction . . . . .	58
3.4 Application to the BRAiNS Study . . . . .	60
3.5 Discussion . . . . .	76
Chapter 4 Discussions and Future Research . . . . .	79
Appendices . . . . .	84
I. Basic quantities in competing risks modeling . . . . .	84
II. A preliminary draft of R code for computing NPMLs . . . . .	85
III. SAS Code for parametric competing risks modeling . . . . .	89
IV. R code for Markov model using “msm” package . . . . .	90
V. R code for running Bayesian inference through MCMC using OpenBUGS . . . . .	91
Bibliography . . . . .	95
Vita . . . . .	99



## LIST OF FIGURES

1.1	Observed times . . . . .	7
1.2	Illness model . . . . .	11
2.1	95% Point-wise confidence interval of CIFs and true CIF for event 1 . . . . .	37
2.2	Cumulative incidence curves of dementia (solid) and death before dementia (dashed) . . . . .	43
2.3	Cumulative incidence function by event and by APOE4 value . . . . .	44
2.4	Cumulative incidence curves of death (black) and MCI (red) . . . . .	48
3.1	Four-state graphical model of BRAiNS study . . . . .	52
3.2	Baseline hazards . . . . .	69
3.3	Sampling history of coefficient for family history . . . . .	75
3.4	Autocorrelation of coefficient for family history . . . . .	76

## LIST OF TABLES

2.1	Coverage probabilities for 95% confidence interval (p95), 90% confidence interval (p90), and 80% confidence interval (p80) for parameters under different right censoring levels . . . . .	39
2.2	Bias, standard error (SE) and coverage probability of 95% confidence interval (p95) of estimated parameters . . . . .	41
2.3	P values from Gray’s tests to compare CIFs among group variables . . .	43
2.4	Regression parameter estimates from Cox proportional hazards (Cox) and Fine and Gray’s proportional SH semi-parametric models . . . . .	45
2.5	Regression parameter estimates from parametric models for cumulative incidence of dementia and death . . . . .	46
3.1	A simple example of states’ record . . . . .	55
3.2	Transition numbers for BRAiNS data . . . . .	61
3.3	Estimated transition intensity matrix . . . . .	61
3.4	Transition probability matrix 5 years after first visit . . . . .	62
3.5	The hazard ratio of each factor for each transition . . . . .	63
3.6	Parameter estimates under time homogeneous hazard rates . . . . .	66
3.7	Baseline hazard rates . . . . .	68
3.8	Parameter estimates under parametric hazard rates . . . . .	70
3.9	Parameter estimate(standard error) from pure semi-Markov model . . . .	72
3.10	Parameter estimate(standard error) from semi-Markov + Markov mixture model . . . . .	73
3.11	Model comparison using DIC . . . . .	74

## Chapter 1 Introduction

### 1.1 Overview

Many research investigations (e.g. Tsiatis et al., 1995; Henderson et al., 2000; Hennerfeind et al., 2006) generate both longitudinal failure time data, with repeated measurements at a number of time points, and event history data, in which time to events of interest are recorded. In such longitudinal failure time data, types of events, corresponding incomplete failure time for each event, and associated covariates are collected from a large number of independent individuals. For example, in the NUN study, 678 subjects were measured annually with the response variable being type of events – “clinical diagnosis”: normal health, mild cognitive impairment (MCI), global impairment, dementia, or death. The other variables include time to first been diagnosed as MCI (or dementia), education, APOE-4 allele (APOE4), family history, etc. The interest of this kind of study is focused primary on the following:

- (a) the distribution of time to competing events;
- (b) the rate and intensity at which each event happens;
- (c) the effects of covariates on conversion and time to events;
- (d) the probability of conversion to dementia before death;
- (e) the effects of the covariates on the above probability.

The event types and event times are our primary response variables. It is common to have incomplete event times, i.e., times are subject to left or/and right censoring, and unknown event types. For example, in the NUN study, some subjects were already demented at the beginning of the study, resulting in left censoring; most event times are only known to be between two consecutive visits, resulting in interval censoring. When the whole longitudinal failure time data is considered, multi-state modeling best describes the problem. On the other hand, classical survival model or compet-

ing risks model work well when only time to events is considered. Survival model is applicable for one event and competing risks model is designed for multiple events. Competing risks models are a sub-category of multi-state models with one starting state, corresponding to one initial state and several absorbing states (e.g. dementia and Death in the NUN study) without any intermediate states. The more general multi-state model is an extension of hazards modeling in survival. For example, the modeling intensities in multi-state is primary based on hazards modeling.

Depending on the questions of interest, either a competing risks model or a more general multi-state model can be constructed. In the following sections, we will introduce competing risks modeling and multi-state modeling, review the previous research on these, present our specific aims, introduce the estimation methods, and finally outline this dissertation.

## 1.2 Competing Risks Models

Time-to-event data with the presence of censored observations arise in studies when there is only one certain event of interest. Kaplan Meier curve, the log rank test, and the Cox proportional hazard model (Cox, 1972) are standard statistical methods for analyzing these kind of data. However, in many situations, there are several distinct events of interest which leads to the competing risks framework. Assume there are  $J$  event types and  $J$  failure times,  $Y_1, \dots, Y_J$ , one for each event. The  $T = \min(Y_1, \dots, Y_J)$  is observed and a variable  $\epsilon = j$  tells which event corresponded to the observed time. So basically,  $(T, \epsilon)$  is observed.

The Cause-specific hazard (CSH) rate for cause  $j$ , is defined by

$$\lambda_j(t) = \lim_{\Delta \rightarrow 0} \frac{P[t \leq T \leq t + \Delta, \epsilon = j | T \geq t]}{\Delta}$$

and the cumulative incidence function (CIF) for  $j$ th competing risk is defined as the following probability

$$F_j(t) = P[T \leq t, J = j] = \int_0^t \lambda_j(\mu) \exp\left\{-\int_0^\mu \sum_{i=1}^J \lambda_i(v) dv\right\} d\mu.$$

Although the fundamental methods for analyzing such data were already developed in the 1970s, researchers are still looking for new methods to approach this type of data. One of the biggest concerns about the competing risk problem is that the associations of dependence among different risks, cannot be identified. As Tsiatis (1975) pointed out, suppose that the set of marginal survival functions are given for some model with dependent risks. Then there exists a unique proxy model with independent risks yielding identical marginal survival functions. There is no way to distinguish the dependent-risk model from the independent risk model when relying solely on the observed data. Thus, it is an important practical aspect that, one can use standard survival methods to analyze cause specific hazard functions, assuming independent risks and regarding all failures due to other causes as additional censoring events. More specifically, suppose we have  $n$  observations and  $J$  risks with  $(t_i, d_i, j_i, x_i)$ , where  $i$  indexes subject,  $x_i$  represents covariate(s) for subject  $i$ ,  $d_i$  is the censoring indicator regarding all risks, and  $\lambda_{j_i}(t_i, x_i)$  is the covariate dependent CSH for event  $j_i$ . Under non-informative censoring from competing risks, the likelihood function can be constructed as

$$L = \prod_{i=1}^n [\lambda_{j_i}(t_i, x_i)]^{d_i} S(t_i, x_i).$$

where  $S(t_i, x_i)$  is the probability of experiencing no events at time  $t_i$ . In terms of hazards,

$$L = \prod_{i=1}^n [\lambda_{j_i}(t_i, x_i)]^{d_i} \prod_{j=1}^k \exp\left\{-\int_0^{t_i} \lambda_j(\mu, x_i) d\mu\right\}.$$

Let  $d_{ji}$  be the indicator for subject  $i$  experiencing risk  $j$ , then

$$L = \prod_{i=1}^n \prod_{j=1}^k [\lambda_j(t_i, x_i)]^{d_{ji}} \exp\left\{-\int_0^{t_i} \lambda_j(\mu, x_i) d\mu\right\}.$$

The likelihood factors as a product of  $k$  likelihoods, one for each type of risk. The likelihood for each type of risk is exactly the same as the one we would obtain by treating all other competing events as censored observations. Therefore, under the assumption of independent risks, the usual Nelson-Aalen estimator can be applied to estimate the cumulative cause-specific hazard. Log rank tests can be used to compare two or more groups. Cox proportional hazard model can be performed for regression analysis. Standard software works sufficiently for these methods, as the competing risks formulation does not add any additional complexity to the analysis. The analysis is valid and useful, but it does not address any correlations among risks, for example, some risks may prevent the subsequent occurrence of others. These methods are limited to CSH functions only and quantities of direct interest may not be included. Kaplan-Meier estimate is another important quantity in survival analysis, which is 1 minus CIF. However, this generally doesn't hold for the cause specific Kaplan-Meier estimate in the competing risks framework. CIF is the more interesting quantity in competing risks.

Competing risks data are typically summarized either by an estimate of hazard function or by direct estimation of an appropriate competing risks probability. Similarly the regression study of the competing risks data falls into two broad categories, modeling the hazard function and direct modeling of the CIF. Compared to the CSH functions, CIFs are considered to be user unfriendly due to a lack of standard statistical software. However, They are straight forward to interpret if we are interested in survival probabilities (or how many subjects fail) for any particular failure type at some time points. Besides this, a CIF is easier to visualize (i.e. graphical display) while a CSH function requires smoothing techniques to achieve better visualization. With these considerations, CIFs should be accompanied by CSH functions in order to better study the competing risks data. The choice is determined by the research question of interest.

It is possible to calculate the CIF from Cox regression by combining estimates from a regression model of all the CSHs (use baseline hazards to calculate a product limit estimate and then CIF). However, the resulting CIF is a complicated nonlinear function of the covariates and difficult to interpret. Also many authors have noted that for a particular failure type, the effects of covariates on the CIF may be very different from those on the corresponding CSH function (Gray 1988, Pepe 1991). Thus, modeling CIF becomes essential in studying competing risks data.

Notice also that the non-identifiability aspect of competing risks models is overturned when covariates are involved, which secures identification of general non-independent competing risks models (Heckman and Honore, 1989). Fine and Gray (1999) introduced direct regression modeling of the CIF, without assuming dependence among risks. In this work, they also describe how the cause-specific regression formulation can be turned into a regression model for CIF via a complimentary log-log transformation. However, most of this work assumes semi-parametric models with one baseline for each risk, which leads to difficulty in the presence of the right censoring. Also, the partial likelihood principle (for one event) no longer works in this case (more than one event). Fine and Gray (1999) and Fine (2001) adapt the inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky 1992) to construct the unbiased estimating function from the complete data partial likelihood. Klein and Andersen (2005) and Andersen et al (2003) proposed an alternative method to handle the censored data, based on pseudovalues from a jackknife statistic constructed from the CIF. These pseudovalues are then used in a generalized estimating equation to obtain estimates of model parameters.

When it comes to inference, the log rank test lacks power if the assumption of proportional hazards among the groups is heavily violated. Therefore, Gray (1988) and Pepe (1991) developed tests for equality of the CIFs across treatment groups. Their methods are useful but restricted to data with grouped variables. In order to quantify

the difference between CIFs, Fine and Gray (1999) suggest to attach to the CIF a kind of hazard function, in the same way we can attach a hazard function to a survival function, which allows us to formulate a proportional hazards model. However, the hazard function defined in this way is different from the usual hazards and lacks an easy interpretation.

Keiding and Andersen (1989) gave explicit formula for computing pointwise confidence intervals for CIFs. A method for simultaneous confidence bands of hazards functions has been suggested by Gilbert et al (2002) with one event and they stated it can be extended to competing risks data. Fine and Gray (1999) also provided a procedure for calculating pointwise confidence intervals and simultaneous confidence intervals (confidence band).

Alternatively, one can try to transform the CIF to another scale and use some type of more generalized linear models on the transformed scale. Jeong and Fine (2006) introduce parametric modeling of CIFs. The parametric method is valuable because compared to nonparametric procedures, it (1) provides better graphical display, even CSH functions have good visualizations; (2) predicts future behavior, assuming the model fits reasonably well; and (3) can assume a specific dependence structure among different event times (Jeong and Fine, 2006).

Typically, in competing risks data, the measurement for a subject who already experienced the first event is stopped (or not included), although some of the other events may still happen afterwards. If this is the case and the follow-up events are also recorded, it causes a problem (Klein,2010) related to the competing risks which is the so-called semi-competing risks problem when the event time for one of the competing risks is known regardless of whether the other risks has occurred or not. In the NUN study, the death time is known for demented subjects but clearly for subjects who died before dementia no dementia time is known. In this case, treating time to dementia as right censored observation when time to death is of interest, as



in CSH modeling, clearly introduces bias and it underestimates the time to death. This problem is more severe when death and MCI are the risks since the time from MCI to death is not ignorable.

Left and interval censoring in competing risks is a more complicated form of the usual competing risks data, which includes only right-censored data. Various estimation methods have been proposed to handle such problem in survival analysis (single event), however, to our knowledge, no prior work has been done on this issue in competing risks data. The fundamental methods to analyze competing risks data in the presence of right censored observations were already developed. However, it is very likely that some subjects encounter an event before registering in the study. For example, in the NUN study, there are more than 100 demented subjects at baseline visit. In order to include these data in the analysis, in this study, we consider the situation where event times are censored also from the left.

In the analysis of competing risks data, it is very often useful to summarize baseline estimates by using a non-parametric likelihood function. Suppose there are  $N$  subjects,  $J$  event types and  $k$  distinct event times:  $t_1 < t_2 < \dots < t_k$ . Suppose  $e_{ji}$  subjects fail from event type  $j$  at time  $t_i$ ,  $c_i$  subjects are right censored in the interval  $[t_i, t_{i+1})$  at times  $t_{i1}, \dots, t_{ic_i}$ , and  $m_i$  subjects (with known event type) are left censored in the interval  $[t_i, t_{i+1})$  at times  $x_{i1}, \dots, x_{im_i}$  (See Figure 1.1). Some basic quantities, which will be used in constructing the likelihood function, are defined in appendix I. The contributions to the likelihood consist of three types for terms:  $F_j(t) - F_j(t-)$

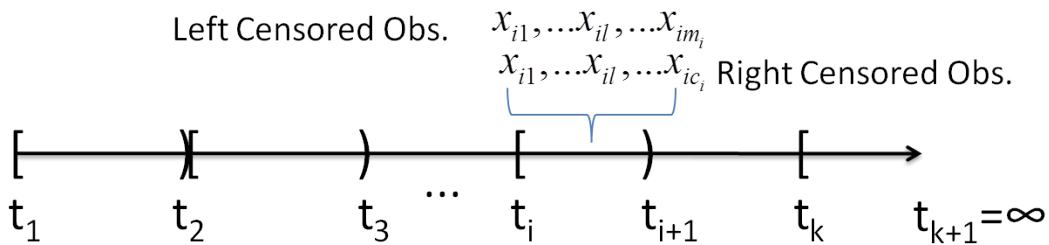


Figure 1.1: Observed times

for observed event type  $j$ ,  $S(t_{il})$  for right censored, and  $F_j(x_{il})$  for left censored event type  $j$ . Thus, the likelihood function can be written as

$$L = \prod_{i=1}^k \left\{ \prod_{j=1}^J [F_j(t_i) - F_j(t_{i-})]^{e_{ji}} \prod_{l=1}^{c_i} [S(t_{il})] \prod_{l=1}^{m_i} \prod_{j=1}^J [F_j(x_{il})]^{e_{jil}} \right\}$$

Note that since  $S(t_{il}) \leq S(t_i)$  and  $F_j(x_{il}) \leq F_j(t_{i+1})$ , the likelihood takes partially maximized values when  $S(t_{il}) = S(t_i)$  and  $F_j(x_{il}) = F_j(t_{i+1})$ , where  $S(t)$  is right continuous and  $F_j(t)$  is left continuous. The partially maximized likelihood can be written as

$$L = \prod_{i=1}^k \left\{ \prod_{j=1}^J [F_j(t_i) - F_j(t_{i-})]^{e_{ji}} [S(t_i)]^{c_i} \prod_{j=1}^J [F_j(t_{i+1})]^{m_i(j)} \right\},$$

where  $m_i(j)$  is the number of left censored observation in the interval  $[t_i, t_{i+1})$  for event type  $j$ . In terms of CSH, as defined earlier,

$$L = \prod_{i=1}^k \left\{ \prod_{j=1}^J (\lambda_{ji}) \prod_{l=1}^{i-1} (1 - \lambda_l)^{e_{ji}} \left( \prod_{l=1}^i (1 - \lambda_l) \right)^{c_i} \prod_{j=1}^J \prod_{l=1}^{i+1} \lambda_{jl} \prod_{q=1}^l (1 - \lambda_q)^{m_i(j)} \right\}$$

where  $\lambda_i = \sum_{j=1}^J \lambda_{ji}$

or

$$L = \prod_{i=1}^k \left\{ \prod_{j=1}^J (\lambda_{ji})^{e_{ji}} \left( \prod_{l=1}^{i-1} (1 - \lambda_l) \right)^{d_i} \left( \prod_{l=1}^i (1 - \lambda_l) \right)^{c_i} \prod_{j=1}^J \prod_{l=1}^{i+1} \lambda_{jl} \prod_{q=1}^l (1 - \lambda_q)^{m_i(j)} \right\}$$

where  $d_i = \sum_{j=1}^J e_{ji}$

There is no direct solution to maximize the above likelihood function. However, under right censoring only, the likelihood simplifies to

$$L^* = \prod_{i=1}^k \left\{ \prod_{j=1}^J (\lambda_{ji})^{e_{ji}} \left( \prod_{l=1}^{i-1} (1 - \lambda_l) \right)^{d_i} \left( \prod_{l=1}^i (1 - \lambda_l) \right)^{c_i} \right\}$$

Let  $n_i$  denote the number of items at risk at a time just prior to  $t_i$ .  $L^*$  can be rewritten as

$$L^* = \prod_{i=1}^k \left\{ \prod_{j=1}^J (\lambda_{ji})^{e_{ji}} (1 - \lambda_i)^{n_i - d_i} \right\}$$

Maximization of the above multinomial likelihood gives the non-parametric maximum likelihood estimates:  $\hat{\lambda}_i = \sum_{j=1}^J \frac{e_{ji}}{n_i}$  and  $\hat{\lambda}_{ji} = \frac{e_{ji}}{n_i}$ ,  $F_j^{\hat{}}(t) = \sum_{t_i \leq t} \frac{e_{ji}}{n_i} F(t_i^-)$ , where  $S^{\hat{}}(t) = \prod_{t_i \leq t} (1 - \hat{\lambda}_i)$ .

Since the number of subjects at risk at time  $t_i$  is unknown in the presence of the left censored data, the above maximization is no longer working. In survival analysis, Turnbull(1976) proposed a computationally based method to find the nonparametric survival function (known as Kaplan-Meier estimator, which can be explicitly formulated for complete or right censored data) for the incomplete data with both left and right censoring. The method is really based on an Expectation-Maximization (EM) algorithm. Similarly, in order to take the left censored data into account in competing risks framework, the iterative EM algorithm will be used in Chapter 2 to obtain the estimates of CIFs.

Investigating covariate effects is our focus as well. Large amounts of left censored observations should be included to increase the sample size and improve the estimation. We modify Jeong and Fine's proportional SH parametric model such that the sum of marginal probabilities (of events) never exceed 1, conditional on any covariates. The left censored data will be accommodated easily using a parametric method. Potential candidates for the parametric probability distribution, which determines the log baseline cumulative SH function, will be examined as well. Simulation studies will be used to investigate the effect of including the left censoring data on the estimates, i.e. variance of estimates etc. Application to investigate the effect of genetics and education on the occurrence of dementia before death in the Nun Study is used to illustrate these results in Chapter 2.

### 1.3 Multi-state Models

A multi-state process is defined as a stochastic process  $(X(t), t \in T)$  with a finite state space  $S = \{0, 1, \dots, s\}$ . Here  $T$  is time interval  $[0, \tau]$ ,  $\tau < \infty$ , and  $X(t)$

takes the value of the state occupied at time  $t$ . The states are either transient or absorbing. An absorbing state is a state from which further transitions cannot occur while a transient state is a state that is not an absorbing state. The multi-state process evolves over time.

Multi-state models are the most common models for describing the longitudinal failure time data. The simplest multi-state model is a two-state model for survival data with states alive (transient) and dead (absorbing). Note that a multi-state with one transient state and multiple absorbing states corresponds to a competing risks model. The illness-death model, with two transient states and one absorbing state, is one of the most important and studied multi-state models. In more complicated models, there can be multiple transient states and multiple absorbing states.

The multi-state process is fully described by transition intensities

$$\alpha_{hj}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(t, t + \Delta t)}{\Delta t}$$

where  $h, j \in S$ ,  $t \in T$ , and

$$P_{hj}(s, t) = pr\{X(t) = j | X(s) = h, \mathcal{F}_{s-}\}$$

are the transition probabilities. Here  $s \leq t$  with  $s, t \in T$  and  $\mathcal{F}_{t-}$  is a  $\sigma$ -algebra generated by history of process over  $[0, t)$ , including state visited, time of transitions, etc.

The full statistical model includes the state structure, which specifies the states and possible transitions and the form of the intensity function for each possible transition. It can be graphically displayed. Figure 2 shows the structure of the illness-death model with transition intensities. Different model assumptions can be made about the dependence of the transition rates on time, resulting in different types of models: (1). Markov models: the transition intensities at time  $t$  do not depend on other aspects of the past history, than the state occupied at  $t$ - (and time-fixed covariates). If the intensities are constant over time, the models are called time homogeneous

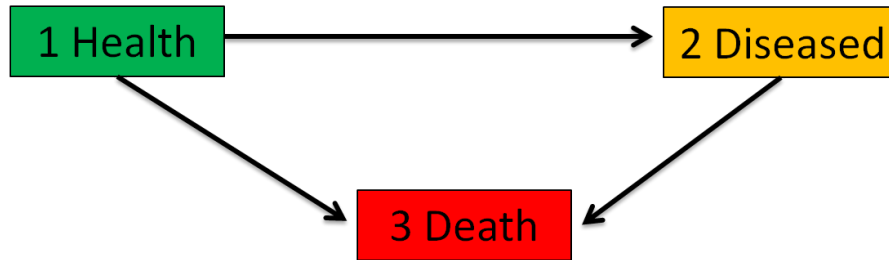


Figure 1.2: Illness model

Markov models. However, often transition probabilities do not just depend on the time between observations, but also on the time origin, resulting in a nonhomogeneous Markov process.

(2). Non-Markov models: besides the state occupied at  $t^-$  (and time-fixed covariates), the transition intensities at time  $t$  depend on other aspects of the past history. If the future evolution depends on the current state and the duration time at current state, the model is called semi-Markov model. If there is only duration dependence, the process is called homogeneous semi-Markov.

Transition intensities, transition probabilities, and state occupation probabilities are the most important quantities in multi-state models. Similar to CSH and CIF in competing risks, both inferences for intensities and probabilities have their advantages. Inference for intensities is simpler and has many standard statistical software support. An attractive feature of multi-state models based on intensities is that all hazard-based models known from survival analysis apply (Andersen et al. 1993). However, it is also of considerable interest to model and to conduct inference for transition probabilities and state occupation probabilities since:

(1) The interpretation of probabilities is simpler and more direct than that of intensities. (For Markov processes, explicit formulas relate such probabilities to transition intensities, thereby allowing for simple plug-in probability estimation once intensity models are established.)

(2) For regression situations, however, plug-in methods do not provide us with simple

parameters describing the association between covariates and outcome probabilities. This is because of the non-linearity of the relation between intensities and probabilities and, hence, even intensity models with a simple link function (such as the Cox model or the additive hazard model), lead to complicated relations between covariates and outcome probabilities. The effect of a covariate on the intensities may be very different from the effect of the covariate on occupation probabilities. For these reasons, both approaches are studied in the literature.

### **Modeling intensities**

Most regression models for intensities relate to the semi-parametric Cox regression model

$$\alpha(t|Z_i) = \alpha_0(t)\exp(\beta^T Z_i)$$

Here  $\alpha_0(t)$  is the baseline hazard function and is left unspecified while the  $\beta$  coefficient are constant, leading to a proportional hazards model. When  $\alpha_0(t)$  is also parametrically modeled, e.g. constant or piecewise constant, the above model turns into parametric hazard model. If the baseline hazard is a function of  $t$  and no other functions of the past is included as a time dependent covariate, the regression leads to a Markov process. The covariates, however, may be time dependent, and the model becomes a non-Markov process.

The Markov assumption has been made in many earlier applications. If time homogeneity is also assumed, the inferences for transition intensities when transition times are known are well studied. Both maximum likelihood estimators and the large sample properties for these estimators have been developed (Albert, 1962). However, the problem with the longitudinal failure time data is that the process is often observed intermittently, which means exact transition times are unknown. The Maximum likelihood estimates (MLEs) can still be evaluated based on numerical maximization (Kalbfleisch and Lawless, 1985). R. C. Gentleman et al. (1993) applied general estimation methods of Kalbfleisch and Lawless, which was for fitting time homogeneous

Markov models originally, to incomplete data due to (i) only a portion of a subject's disease history is observed, (ii) interval censored times because of intermittent visits, (iii) unknown disease onset time, with arbitrary transition structure. They also provided methods for assessing the modeling of incomplete data. The large sample properties for these MLEs have been developed by Bladt and Sorensen (2005).

Non-homogeneous Markov models are studied when the time homogeneity assumption is violated. Joly et al. (2002) show that in a nonhomogeneous Markov model for an illness-death model with interval-censored data, estimating the transition intensity from health to illness by survival analysis (treating death as a censored event) is biased downwards. They proposed a method to get smooth estimates of the transition intensities,  $\hat{\alpha}_{01}$ ,  $\hat{\alpha}_{12}$  and  $\hat{\alpha}_{02}$ , by maximizing a penalized likelihood.

$$pl(\alpha_{01}, \alpha_{12}, \alpha_{02}) = l(\alpha_{01}, \alpha_{12}, \alpha_{02}) - k_{01} \int \alpha''_{01}(\mu) d\mu - k_{12} \int \alpha''_{12}(\mu) d\mu - k_{02} \int \alpha''_{02}(\mu) d\mu$$

where  $l$  is the full log-likelihood,  $\alpha_{hj}(\mu)$  are intensities and  $k_{hj}$  are three positive smoothing parameters. Joly and Commenges (1999) suggested that the penalized likelihood can be used in a regression model, with regression parameters and baseline transition intensities being estimated simultaneously.

Alternatively, Hubbard et al. (2008) decomposed the nonhomogeneous transition intensity matrix into the product of a baseline transition intensity matrix and a scalar function of time, and proposed to deal with nonhomogeneous Markov process via a time transformation. The time scale of the nonhomogeneous Markov process is transformed to an operational time scale on which the process is homogeneous. Time transformation and parameters for homogeneous Markov process were then jointly modeled.

The most important deviation from the Markov property in practice is duration dependence, resulting in a Semi-Markov model. Semi-Markov models have been considered in some other applications. Foucher et al. (2010) present a flexible semi-Markov model for interval censored data. The model is a combination of Markov chain

and distribution of durations. The explanatory variables were introduced through Markov chain and through probability density functions of durations (generalized Weibull distribution).

### Modeling probabilities

In Markov models,

$$P_{hj}(s, t) = pr\{X(t) = j | X(s) = h\},$$

combines both direct and indirect transitions from state h to state j. And the transition probability matrix can be calculated from transition intensity matrix by means of a product integral (Aalen and Johansen, 1978)

$$P(s, t) = \prod_{(s,t]} (I + dA(\mu))$$

More generally, for Markov models with time-fixed covariates,  $Z_0$ ,  $P_{hj}(s, t | Z_0)$  can be estimated completely analogously by the product integral. Variance estimates may be obtained from those of intensities via delta-method (Shu and Klein, 2005). Explicit expressions for transition probabilities in an integral form are available for semi-Markov processes without loops (or non-reversible paths) (Andersen and Perme, 2008). However, again, the relations between covariates and transition probabilities are complicated and there emerges modeling probabilities directly.

Andersen et al. (2003) present a technique which models the state probability  $Q_{hi}(t_k) = P(X_i(t_k) = h)$  directly through a link function  $g()$ :

$$g(Q_{hi}(t_k)) = \alpha_t + \beta^T Z_i$$

where  $Z$  are the covariates.

The link function  $g()$  can be logit, probit or complementary log-log function. The regression was based on state probabilities obtained from Aalen-Johansen estimator. The method uses the pseudo-observations from a jackknife statistic constructed from simple summary statistic estimates of the state probabilities,

$$\hat{Q}_{hi}(t) = \hat{Q}_h(t) - (n - 1)\hat{Q}_h^{-i}(t)$$



where  $\hat{Q}_h(t)$  is based on the entire sample and  $\hat{Q}_h^{-i}(t)$  is the same estimator but eliminating subject  $i$ , the idea being that the leave-one-out diagnostics for the summary statistic contain information about the way in which covariates for each individual affect the estimator.

These pseudo-values are then used in a generalized estimating equation to obtain estimates of the model parameters.

$$U(\beta) = \sum_i U_i(\beta) = \sum_i \left( \frac{\partial}{\partial \beta} g^{-1}(\beta^T Z_i) \right)^T V_i^{-1} (\hat{\theta}_i - g^{-1}(\beta^T Z_i)) = 0.$$

Where  $\theta_i = (\hat{Q}_{hi}(t))$ . The questions left open are the choice of time points and working covariance matrix in the generalized estimating equation.

The subject specific random effects models, in particular, generalized linear mixed model (GLMM), take the heterogeneity across subjects into account, in addition to the fixed effects. Salazar (2004) and Salazar et al. (2007) proposed a discrete-time multi-state Markov model with shared random effects, which assumes that the follow-up response component and drop-out response component are sharing the same random parameters and are conditionally independent given these random effects, for longitudinal data with categorical responses. The transition probabilities were modeled through GLMM structure (multinomial logistic model). The great advantage of the GLMM structure is that inference and regression can be done through standard statistical software. However, the joint distribution of the response variable was a conditional distribution given the baseline information, resulting in possibly a so-called “baseline confounding” problem, especially when the baseline (or initial) states are different for different subjects. Yu et al. (2009) continued to model using shared random effects approaches by incorporating the baseline distribution into the followup likelihood. All these studies applied the models to the NUN study with three transient states (intact cognition, mild cognitive impairment(MCI), and global impairment) and two competing absorbing states (death and dementia). However,

the parameters resulting from these models on transition probabilities do not have straightforward interpretation.

Time-homogeneous Markov process is not appropriate in many real data, for example, the transition intensity of death from health is clearly age dependent. Thus a model with time dependent transition intensities is usually required. We consider both constant transition intensities and duration time dependent transition intensities, leading to a mixture of Markov and semi-Markov processes, and model both transition intensities, and transition probabilities. The model allows interval censored data and could incorporate the random effects, accounting for the unobserved subject heterogeneity. The likelihood for the multi-state model is constructed through a multinomial distribution. In Chapter 3, we show the covariate effects on transition intensities, overall transition probabilities, and duration times in the BRAiNS data.

## 1.4 Estimation Methods

### 1.4.1 Maximum likelihood estimation (MLE)

Based on the data, a model is defined as the family of probability distributions indexed by unknown parameters  $p(x|\Theta)$ . Assuming independent observations,  $p(x|\Theta) = p(x_1, \dots, x_n|\Theta) = \prod_{i=1}^n p(x_i|\Theta)$  and the likelihood has the same expression but written as  $L(\Theta|x)$ . Given that different parameter values index different probability distributions, our desired probability distribution is the one that makes the observed data most likely: the one that maximizes the likelihood function  $L(\Theta|x)$ , resulting in MLE. MLE is a standard approach that can be developed for a large variety of parameter estimation and inference situations. MLE has many optimal properties, such as sufficiency, consistency, efficiency, and invariance.

Usually the MLE is maximizing the log-likelihood function (or minimizing the nega-

tive log-likelihood function) and obtained by solving the likelihood equation:

$$\frac{\partial \log L(\Theta|x)}{\partial \theta_i} = 0$$

and checking the following:

$$\frac{\partial^2 \log L(\Theta|x)}{\partial \theta_i^2} < 0.$$

However, in practice, it is usually difficult to obtain such an analytic solution, especially when non-linear equations needed to be solved. There are numerical optimization algorithms sought to maximize the log-likelihood: Newton-Raphson Method, Quasi-Newton Methods, Conjugate Gradient Methods, and so on. They differ in how the updating routine for searching improved parameters is conducted. We take Newton-Raphson algorithm for example. The basic Newton-Raphson algorithm can be described as:

Suppose  $f(x)$  is the function to be maximized, one starts with a starting value  $x_0$  which is reasonably close to maximize the function. Then the function is approximated by a “quadratic approximation” based on a Taylor series expansion of the function.

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$

To maximize a quadratic function, take derivative and set equal to zero, and then solve the linear equation we get

$$x = x_0 - \frac{f'(x_0)}{f''(x_0)}$$

This new  $x$  will typically be a better approximation than the starting value, and the method can be iterated.

The R package 'optim' can be used to find the minimum of the negative log-likelihood, in which an approximate covariance matrix for the parameters is obtained by inverting the Hessian matrix at the optimum. Alternatively, The NLP procedure (NonLinear Programming) in SAS offers a set of optimization techniques for minimizing or maximizing a continuous nonlinear function subject to linear and nonlinear, equality and

inequality, and lower and upper bound constraints. Problems of this type are found in many settings ranging from optimal control to maximum likelihood estimation (SAS online documentation). PROC NLP implements many optimization algorithms. These optimization techniques require a continuous objective function and some of them also require continuous second-order derivatives. There are three ways to compute derivatives in PROC NLP: analytically (using a special derivative compiler), finite difference approximations, or user-supplied exact or approximate numerical functions. The factors that go into choosing a particular optimizer for a particular problem are complex, but SAS documentation provides some guidance for this.

MLE by these optimization algorithms can be sensitive to the choice of starting values and it is worth noting that the optimization algorithm does not necessarily guarantee that true maximization will be achieved. Finding optimum parameters is essentially a heuristic process in which the optimization algorithm tries to improve upon an initial set of parameters that is supplied by the user. One way to verify a global maximum is choosing different starting values over multiple runs of the iteration procedure and examining the results to see whether the same solution is obtained repeatedly. Since the parameters for SH are comparable with those obtained by Cox PH models, parameter estimates resulting from Cox models will be used as initial values to obtain estimates in our model.

#### **1.4.2 Bayesian inference**

Typical statistical inferences involve estimating parameters, given the available data. The classical frequentist approach assumes that parameters are fixed and data are a repeatable random sample. MLE generally provides the solution in this approach. The Bayesian approach assumes that data are fixed and parameters have an unknown joint distribution. This method begins with some prior distribution and updates pos-

terior distribution based on the observed data. Unfortunately, most realistic posterior distributions are high dimensional, especially when a random effect is considered, and analytic solutions are hard to obtain. Point estimates and uncertainty intervals are often used for inference.

Bayesian approach requires both a likelihood function based on data  $\underline{X}$  and parameters  $\Theta$ , and a prior probability distribution for these parameters. With the likelihood and the prior, Bayes' formula

$$p(\Theta|\underline{X}) = \frac{p(\underline{X}|\Theta) * p(\Theta)}{\int_{\Theta} p(\underline{X}|\Theta) * p(\Theta) d\Theta}$$

gives a posterior distribution for the parameters, and all Bayesian inferences are based on this.

There are three basic terms in Bayesian inference: point estimation, interval estimation, and hypothesis testing. The simplest is point estimation which is usually estimated by the posterior mean or mode. A Bayesian interval estimate is called a credible interval. The hypothesis testing in the Bayesian approach is based on calculating the posterior probability that the null hypothesis is true and the decision is made depending on the value of this posterior probability.

After subjectively specifying the prior distribution, which draws criticism of the Bayesian approach, the critical step in Bayesian approaches is computing the posterior distribution. This often requires the integration of high-dimensional functions and the solutions are obtained through numerical techniques. The most widely applicable numerical technique for sampling from the posterior is Markov Chain Monte Carlo(MCMC) using the Gibbs sampler. It consists of Monte Carlo integration, Markov chain and Gibbs sampler.

### **Monte Carlo integration**

The original Monte Carlo approach was a method developed by physicists to use random number generation to compute integrals. Suppose we want to compute the

integral

$$\int f(x)dx = \int g(x) * p(x)dx,$$

where  $p(x)$  is a density function and the integral becomes the expectation of  $g(x)$ . Consider  $N$  random points from  $p(x)$  distribution:  $x_1, \dots, x_N$ , evaluate the function  $g(x)$  at each point and take the average to obtain the expectation and thus the integral.

$$\int f(x)dx \approx \frac{1}{N} \sum_1^N g(x_i)$$

### Markov Chain

In MCMC, the current sample values are always based on the previous sample values and transition probabilities between sample values are only a function of the most recent sample value, resulting in a Markov chain.

Let  $X(t) \in S$  denote the value of a random variable at time  $t$ , where  $S$  is the state space. Suppose the process is a Markov process and  $P$  is the transition matrix, then we have

$$\pi_{t+1} = \pi_t P,$$

where  $\pi(t)$  is the distribution of  $X$  at time  $t$ .

Under some conditions, i.e., the chain is irreducible and aperiodic, there exists a stationary distribution  $\pi^*$  such that

$$\pi^* = \pi^* P.$$

The above basic idea of discrete-state Markov chain can be generalized to a continuous state Markov process by having a probability kernel  $P(x; y)$  satisfying

$$\int P(x; y)dy = 1.$$

The Chapman-Kolmogorov equation in this continuous case becomes

$$\pi_t(y) = \int \pi_{t-1}(x)P(x; y)dy,$$

where  $\pi_t(y)$  is the distribution at time  $t$ . The stationary distribution  $\pi(y)$  satisfies

$$\pi(y) = \int \pi(x)P(x; y)dy.$$

### Gibbs sampling

The remaining part comes to random sampling from a complex probability distribution  $p(x)$ . One of the important approaches is Metropolis-Hastings sampling. Suppose our goal is to generate samples from some distribution  $p(\theta)$ . The Metropolis-Hastings algorithm proceeds as follows:

1. Start with an initial value of  $\theta_0$ .
2. Using the current state value  $\theta_t$ , draw a candidate state value  $\theta_{can} \sim q(\theta_t, \theta)$ , a proposed density which gives the probability of returning a value  $\theta$  given a previous value of  $\theta_t$ . Note the original Metropolis algorithm requires  $q(\theta_t, \theta) = q(\theta, \theta_t)$  and Hastings lifts the restriction, allowing the proposed density to be independent of the current state value.
3. Let  $\theta_{t+1} = \theta_{can}$  with probability

$$\alpha = \min\left(\frac{p(\theta_{can})q(\theta_{can}, \theta_t)}{p(\theta_t)q(\theta_t, \theta_{can})}, 1\right).$$

4. Repeat steps 2 and 3.

It was proved that the Metropolis-Hasting sampling generates a Markov chain whose stationary distribution is the candidate density  $p(\theta)$ .

Gibbs sampling is a special case of the Metropolis-Hastings sampling with  $\alpha = 1$ . It is applicable when the joint distribution is not known explicitly or is difficult to sample from directly, but the conditional distribution of each variable is known or is easier to sample from. The Gibbs sampling algorithm generates an instance from the distribution of each variable in turn, conditional on the current values of the other variables. Thus, one simulates  $k$  random variables sequentially from the  $k$  conditionals rather than generating a single  $k$ -dimensional vector using the full joint distribution. More specifically, let  $(\theta_1, \dots, \theta_k)$  denote the parameters. The sampler

works as follows:

1. Start with initial values  $(\theta_1^{(0)}, \dots, \theta_k^{(0)})$ .
2. Given the current state value of the chain  $(\theta_1^{(t)}, \dots, \theta_k^{(t)})$ , generate the next state according to:

$$\begin{aligned} \theta_1^{(t+1)} &\sim p(\theta_1 | \theta_2^{(t)}, \dots, \theta_k^{(t)}) \\ \theta_2^{(t+1)} &\sim p(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_k^{(t)}) \\ &\cdot \quad \cdot \\ &\cdot \quad \cdot \\ &\cdot \quad \cdot \\ \theta_k^{(t+1)} &\sim p(\theta_k | \theta_1^{(t+1)}, \dots, \theta_{k-1}^{(t+1)}). \end{aligned}$$

Again the theory implies that the Markov chain is guaranteed to converge to the appropriate stationary distribution.

Bayesian inference is able to handle missing data, especially when the response variable is missing. The method is promising since it avoids the computation of the exact likelihood. For example, let  $\Theta$  denote all regression coefficients and  $\Sigma$  the vector of all variance parameters, then the full Bayesian inference is based on the posterior distribution

$$f(\Theta, \Sigma | data) \propto L(\Theta, \Sigma) f_0(\Theta, \Sigma),$$

where  $L$  is the likelihood and  $f_0$  is the prior distribution. By MCMC using the Gibbs sampler, we generate a sequence of  $N$  samples, i.e.

$$(\theta^{(1)}, \sigma^{(1)}), (\theta^{(2)}, \sigma^{(2)}), \dots, (\theta^{(N)}, \sigma^{(N)}).$$

Now suppose  $T$  is our response variable and  $T$  is missing for some observations. The Bayesian inference with data augmentation basically adds in one more step for each scan and it includes the following two steps:

1. Augmentation (or imputation) step:



With a current regression coefficients  $\theta^{(m)}$  and variance parameters  $\sigma^{(m)}$ , simulates the missing values for each observation independently, i.e. draw values from a conditional distribution

$$p(T_{missing}|\theta^{(m)}, \sigma^{(m)}),$$

where  $T_{missing}$  denotes the missing observations for the response variable.

2. Posterior sampling step:

With the complete sample data, the new posterior regression coefficients  $\theta^{(m+1)}$  and variance parameters  $\sigma^{(m+1)}$  were simulated.

The two steps are iterated long enough for the results to be reliable for a multiply imputed dataset. Let  $m$  range from 1 to  $N$ . The following Markov chain will be created:

$$(\theta^{(1)}, \sigma^{(1)}, T_{missing}^{(1)}), (\theta^{(2)}, \sigma^{(2)}, T_{missing}^{(2)}), \dots, (\theta^{(N)}, \sigma^{(N)}, T_{missing}^{(N)}),$$

which converges to  $p(T_{missing}, \Theta, \Sigma)$  as  $N \rightarrow \infty$ .

The application of Bayesian inference has been greatly facilitated by MCMC, which constructs a Markov chain with stationary distribution equal to the posterior distribution of interest. Once the chain has converged, realizations are considered as dependent samples from the distribution and thinned realizations are considered to be approximate independent samples.

As for applying these algorithms to generate samples, there are many programs available for such computations. We will focus on OpenBUGS(Lunn. 2009), a free software package (<http://www.openbugs.info/w/>), that is the most popular and has good documentation. With BUGS, one simply needs to make some general specifications about the model and provide some initial values. A typical OpenBUGS proceeds as follows:

- (a).Specify the model to run and prior distributions for all parameters to be estimated. This step basically constructs the likelihood.
- (b).Provide data and initial values.

(c).Generate the MCMC simulations after checking the model, data, and initial values.

There is a list of standard distributions and therefore a list of standard models provided by OpenBUGS. However, we may use the “zeros trick” (OpenBUGS User Manual) for non-standard distributions. Suppose we wish to use a sampling distribution that is not included in the list of standard distributions provided by OpenBUGS, in which an observation  $x_i$  contributes a likelihood term  $L_i$ . Recall a  $\text{Poisson}(\phi)$  with observation of zero has likelihood  $\exp(-\phi)$ , so if our observed data is a set of 0’s, and  $\phi_i$  is set to  $-\log(L_i)$ , we will obtain the correct likelihood contribution. (Note that  $\phi_i$  should always be  $> 0$  as it is a Poisson mean, and so we may need to add a large enough positive constant to ensure that it is positive.) This trick allows arbitrary sampling distributions to be used, and is particularly suitable when dealing with interval censored observations. The trick is illustrated in the following example.

```
C <- 10000
for (i in 1:N) {
  zeros[i] <- 0
  phi[i] <- -log(L[i]) + C
  zeros[i] ~ dpois(phi[i])
}
```

Note that initial samples from each run are not valid because the Markov chain has not stabilized yet. The “burn in samples” strategy is often used, which discards some initial samples. Highly correlated parameters cause high autocorrelation, which is important since high autocorrelation will take a very long time for the simulated samples to explore the entire posterior distribution. Typically, the level of autocorrelation declines with an increasing number of lags in the chain. Therefore, the strategy to reduce the autocorrelation is to “thin” the chain by taking every  $i$ th sam-

ple. However, we need to identify when the stationary distribution is achieved?. One way is to carry out multiple independent runs to see how stable the estimates are. Trace plot, autocorrelation, and Gelman-Rubin are common tools used to diagnose the convergence of the MCMC samples after excluding the initial burn in samples and after thinning. The Gelman-Rubin diagnostics rely on parallel chains to test whether they all converge to the same posterior distribution. The “potential scale reduction factor” is calculated for each variable in  $x$ , together with upper and lower confidence limits. Approximate convergence is diagnosed when the upper limit of “potential scale reduction factor” is close to 1.

In order to accommodate the missing data in MCMC simulation, missing values are treated as unknown parameters and are provided with some initial values. During each step of sample updating in MCMC simulation, these missing values can be imputed based on some proposed models, or even the EM algorithm. OpenBUGS automatically imputed missing response values. Missing covariate values needs to be imputed by specifying models for them. OpenBUGS is incredibly flexible to deal with the missing data issue. Enter an 'NA' in the appropriate matrix cell in OpenBUGS for a data point that is missing. The program will treat all of the missing elements of the data matrix as if they were unknown model parameters. Specify the probability model to those parameters and the program will run the sampling.

## **1.5 Outline of the Dissertation**

The remainder of this dissertation is divided into two major parts: competing risks modeling and multi-state modeling, and a summarization is made over these two. It is organized as follows:

In Chapter 2, we consider Jeong and Fine (2007)'s proportional SH model. We modify their model to assure that the sum of the underlying CIFs never exceeds one.

Simulation studies were conducted to investigate the effect of left censored data on the parametric estimation. Then, the proposed model was applied to investigate the effect of genetics and education on the occurrence of dementia before death in the Nun Study.

In Chapter 3, We propose a model with a mixture of Markov and semi-Markov process, where both transition intensities and transition probabilities are modeled. The likelihood for the multi-state model is constructed through a multinomial distribution. The model allows interval censored data and could incorporate the random effects. The estimation are obtained through Bayesian approach using OpenBUGS. As an application, we show the covariate effects on transition intensities, overall transition probabilities, and duration times in the BRAiNS data.

Finally, Chapter 4 summarizes our work. A discussion of findings is presented along with the conclusions derived. Moreover, potential areas for further study are briefly discussed.

## Chapter 2 Competing Risks Modeling

In the Nun study, death is a competing risk encountered when following a cohort of elderly subjects to dementia. Analyses in medical papers frequently assume independent risks for competing events and evaluate covariate effects on these events by modeling distinct proportional hazards regression models for each event. Jeong and Fine (2007) proposed parametric modeling of CIF without assumptions about dependence among events. We modified their model to assure that the sum of the underlying CIFs never exceeds one, by assuming a proportional SH model for dementia only. 106 out of 268 demented subjects are demented at baseline, which introduces the left censoring data. To our knowledge, no prior work has been done on this issue in competing risks data. We computed non-parametric MLE of CIF based on EM algorithm and investigated the covariate effects by applying our proposed parametric model after including left censored data. After including left censored data, the incidence rate of dementia becomes larger than that of death for age  $< 90$ , education becomes significant factor for incidence of dementia and standard errors for estimates are smaller.

### 2.1 Introduction

Competing risks problems (Klein, 2010) are common in medical studies. For example, in Alzheimer disease research, an elderly person is at risk for multiple events, including dementia and death. In the competing risks problem, the basic quantities are cause specific hazard (CSH) functions and cumulative incidence functions (CIFs), whose value at time  $t$  is the probability of failure by time  $t$  for a particular type of failure in the presence of other risks.

Analyses in medical papers (Lunn and McNeil, 1995) frequently assume independent

risks for competing events and evaluate covariate effects on these events by modeling distinct proportional hazards regression models for each event. Fine and Gray (2001) recommended an improved semi-parametric method for analyzing competing risks data based on CIF without assuming independence. The non-parametric maximum likelihood estimators (MLEs) of CIFs can be analytically calculated and graphically plotted. Covariate effects on CIFs can be estimated and inferred through their semi-parametric regression. Alternatively, Jeong and Fine (2006) shown that one can transform the CIF to another scale using a link function, i.e. complementary log-log function. They then apply a parametric modeling approach which provides better graphical display, predicts future behavior, and does not assume independence among different event times. However, the improper baseline sub-distribution function is critical to the success of this parametric modeling. Jeong and Fine (2007) found that an improper Gompertz distribution is a good approximation to the CIF in their data.

The problem with Fine and Gray's semi-parametric proportional SH model and Jeong and Fine's parametric model on CIFs is that the overall probability of any event occurring could exceed one. To address this problem, we assume a proportional SH model for the event of interest through a complementary log-log link function and assume a similar link function for the other events. This modification ensures the overall probability of any event occurring is one. Such an asymmetric model is useful when we are interested in one specific event. For example, in the Nun Study, we focus on the event of dementia before death and covariate effects on it. We compute MLEs of the covariate effects by assuming the baseline sub-distribution as the product of a Weibull distribution and an unknown constant. The goodness-of-fit of this solution is examined in comparison to non-parametric MLEs of CIFs.

Left censoring is a second and more complicated issue since the non-parametric MLEs can no longer be analytically computed. Instead, the non-parametric MLEs of the

survival function, hazard rates, and CIF are computed based on competing risks data with only right censoring using the Expectation-Maximization (EM) algorithm. The regression analysis of the left censored data will be studied using the same parametric method as described above. Simulation studies will be used to investigate the effect of the left censored observations on the estimates, i.e. bias, standard deviation, and confidence interval coverage.

We apply our proposed methods to analyze data on dementia and death as competing events in the Nun Study, a population based cohort study of aging and Alzheimer's disease. The dataset consists of a cohort of 678 members of the school sisters of Notre Dame religious congregation (Snowdon et al., 1997). Each participant agrees to allow investigators complete access to their convent archives, participate in near-annual assessments of cognitive and physical function, and donate their brain at death. Of these, 72 participants are excluded from the analysis because of missing covariates or consent withdrawal. Final events (first dementia, death, or censored) and covariates form a competing risks data set. Subjects were classified as demented if they met clinical criteria for dementia: impaired on Delayed Word Recall and at least one other cognitive test and impaired on Activities of Daily Living. Among these 606 observations, 74 are right censored, 268 are demented, and 264 experienced death before dementia. Also, 106 out of 268 demented subjects are demented at baseline, resulting in left censored data. Since this is an observational study, we don't account for the potential left truncated data, i.e. individuals who died before being able to register for the study.

In Section 2, we introduce notation and formulas on the proposed models for CIF. In Section 3, the results of several simulation studies are reported. In Section 4, we apply our methods to the Nun Study data. These models are further compared to Fine and Gray's semi-parametric proportional SH model and the naive Cox proportional hazards model. Finally, Section 5 contains a discussion and some conclusions.

## 2.2 Model Setup

Assume there are  $k$  event types and  $k$  failure times,  $Y_1, \dots, Y_k$ , one for each event. Then  $T = \min(Y_1, \dots, Y_k)$  is observed and a variable  $\epsilon = j$  tells which event corresponds to the observed time. Some observations are subject to right censoring by follow-up time, indicated by  $d$ , and some observations are left censored by registration time, indicated by  $l$ . Event types are known for left censored observations. Also observed are covariates  $X$ . So basically,  $(T, \epsilon, d, l, X)$  is observed. Instead of hazard rate and survival function, CSH rate and CIF are the two basic quantities in the competing risks framework (Fine, 2001; Klein and Andersen, 2005; Klein, 2006). The CSH rate for cause  $j$  is defined as  $\lambda_j(t) = \lim_{\Delta \rightarrow 0} \frac{P(t \leq T \leq t + \Delta, \epsilon = j | T \geq t)}{\Delta}$ , and CIF for  $j$  event is  $F_j(t) = P[T \leq t, \epsilon = j] = \int_0^t \lambda_j(\mu) \exp(-\int_0^\mu \sum_{i=1}^k \lambda_i(\nu) d\nu) d\mu$ .

### 2.2.1 Non-parametric MLEs

Suppose there are  $N$  observations and  $n$  distinct observed event times:  $t_1 < t_2 < \dots < t_n$ . Suppose  $d_{ji}$  subjects fail from event type  $j$  at time  $t_i$ . Let  $n_i$  denote the number of subjects at risk at a time just prior to time  $t_i$ ,  $\lambda_i$  denote overall hazard rate and  $\lambda_{ji}$  denote CSH rate for event  $j$  at time  $t_i$ . Suppose we have no left censored observations. Then according to Kalbfleisch and Prentice (2002), the non-parametric MLEs are given by:  $\hat{\lambda}_i = \sum_{j=1}^k \frac{d_{ji}}{n_i}$ ,  $\hat{\lambda}_{ji} = \frac{d_{ji}}{n_i}$ , and CIF for event  $j$ :  $\hat{F}_j = \sum_{t_i \leq t} \frac{d_{ji}}{n_i} \hat{F}(t_i -)$ , where overall survival function  $\hat{F}(t) = \prod_{t_i \leq t} (1 - \hat{\lambda}_i)$ .

In the presence of left censored observations, the number of subjects at risk at time  $t_i$  and the number of subjects experiencing event  $j$  are unknown. As a result, CSH rates and CIF at time  $t_i$  can not be analytically computed. So the above maximization is no longer applicable. We propose an iterative EM algorithm which was first discussed in Turnbull (1974) for estimating survivorship function. Let  $l_{ji}$  be the number of left



censored observations (event  $j$ ) between observed event times  $t_{i-1}$  and  $t_i$ . We use self-consistent estimators (Turnbull, 1974) to estimate the probability that this event occurred at each observed event time, less than  $t_i$ , based on initial estimates of CIFs. Using this estimation, we compute an expected number for each type of event at each observed time  $t_i$  and an expected number of subjects at risk at each observed time  $t_i$ , which are then used to update the estimates of CIFs. This procedure is repeated until CIFs are stabilized.

Given updated event number  $d_{ji}^*$  and updated number of subjects at risk  $n_i^*$ , it is straightforward to compute  $\hat{\lambda}_{ji}^*$ ,  $\hat{\lambda}_i^*$ ,  $\hat{F}^*(t)$ , and  $\hat{F}_j^*(t)$ , i.e.  $\hat{\lambda}_{ji}^* = (d_{ji}^*)/(n_i^*)$ . This fact is important in computing the CIF in the presence of left censoring using the EM iterative algorithm as follows.

Step 1. Choose initial estimates of CIFs at each observed time  $t_i$ ,  $F_j(t_i)$ . Any legitimate estimates will work and good initial estimates are CIFs obtained by ignoring the left censored observations.

Step 2 (E-step). Using the current estimates of CIFs, compute the expected number of events (for each type of event) and expected number of subjects at risk,  $d_{ji}^*$  and  $n_i^*$ , note these numbers can be non-integer.

Step 3 (M-step). With the results from previous step, maximization is performed as if there is only right censoring, yielding  $\hat{\lambda}_{ji}^* = \frac{d_{ji}^*}{n_i^*}$ ,  $\hat{\lambda}_i^* = \sum_{j=1}^k \frac{d_{ji}^*}{n_i^*}$ ,  $\hat{F}^*(t) = \prod_{t_i \leq t} (1 - \hat{\lambda}_i^*)$ , and  $\hat{F}_j^*(t) = \sum_{t_i \leq t} \frac{d_{ji}^*}{n_i^*} \hat{F}^*(t_i^-)$ .

Step 4. Return to step 2 and proceed iteratively until convergence.

## 2.2.2 Semi-parametric models

The regression model for CSH is based on the Cox proportional hazards model (Cox, 1972) and assumes non-informative censoring from competing risks. The CSH for event type  $j$  is a multiplicative function of the baseline hazard  $\lambda_{j0}(t)$ , given a single

covariate  $X$ :

$$\lambda_j(t; X) = \lambda_{j0}(t) \exp(\gamma_j X) \quad (2.1)$$

where  $\exp(\gamma_j X)$  is the hazard ratio for event type  $j$ . A similar form of regression model was proposed by Fine and Gray (1999) by assuming a proportional sub-distribution hazard (SH,  $\alpha$ ) model with

$$\alpha_j(t; X) = \alpha_{j0}(t) \exp(\beta_j X) \quad (2.2)$$

where  $\exp(\beta_j X)$  is the sub-distribution hazard ratio for event type  $j$ , and SH is associated with the CIF:  $\alpha_j(t; X) = \frac{dF_j(t; X)}{1 - F_j(t; X)}$ . The major difference between CSH and SH is that SH is directly related to the CIF.

### 2.2.3 Parametric model

The proportional SH model corresponds to a direct transformation modeling of CIF defined by an invertible and monotonically increasing non-parametric function,  $\mu_j(t)$ , which determines the baseline CIF for event  $j$  when  $X = 0$ :

$$g_j(F_j) = \mu_j(t) + \beta_j X \quad (2.3)$$

where  $g_j(F_j) = \log\{-\log(1 - F_j)\}$  and  $\exp(\beta_j)$  is the sub-distribution hazard ratio for event type  $j$ . Equivalently, the CIF for event  $j$  is given by

$$F_j(t; X) = 1 - \exp(-\exp(\beta_j X) * \exp(\mu_j(t))) \quad (2.4)$$

The direct parametric modeling of CIF was proposed by Jeong and Fine (2007) by specifying a parametric form for  $\mu_j(t)$ . Based on their transformation models, for example with complementary log-log link function, the probability of experiencing a type  $j$  event is

$$F_j(\infty; X) = 1 - \exp(-\exp(\beta_j X) * \exp(\mu_j(\infty))) = 1 - [\exp(-\exp(\mu_j(\infty)))]^{\exp(\beta_j X)} \quad (2.5)$$

where  $P_{j0} \leq 1$ . Now suppose there are two competing events and the largest times are not subject to censoring, i.e.  $P_{10} + P_{20} = 1$ . For  $X \neq 0$ ,  $F_1(\infty; X) + F_2(\infty; X) = 1 - [\exp(-\exp(\mu_1(\infty)))]^{\exp(\beta_1 X)} + 1 - [\exp(-\exp(\mu_2(\infty)))]^{\exp(\beta_2 X)}$ , which generally is not equal to  $P_{10} + P_{20}$  and could possibly be greater than 1, resulting in a problematic modeling. The same problem exists in Fine's semi-parametric proportional SH model, as shown in the following:

Based on the definition for SH, we have

$$\alpha_j(t; X) = \frac{dF_j(t; X)}{1 - F_j(t; X)}$$

$$\int_0^t \alpha_j(\mu; X) d\mu = -\log(1 - F_j(t; X)).$$

Assume there are only two events, then we have

$$F_1(\infty; X) = 1 - \exp\left(-\int_0^\infty \alpha_1(\mu) \exp(\beta_1 X) d\mu\right) = 1 - A^{\exp(\beta_1 X)}$$

and

$$F_2(\infty; X) = 1 - \exp\left(-\int_0^\infty \alpha_2(\mu) \exp(\beta_2 X) d\mu\right) = 1 - B^{\exp(\beta_2 X)}$$

where  $A = \exp(-\int_0^\infty \alpha_1(\mu) d\mu)$  and  $B = \exp(-\int_0^\infty \alpha_2(\mu) d\mu)$ .

Thus,

$$F_1(\infty; X) + F_2(\infty; X) = 1 - A^{\exp(\beta_1 X)} + 1 - B^{\exp(\beta_2 X)}.$$

We know that  $F_1(\infty; X = 0) + F_2(\infty; X = 0) = (1 - A) + (1 - B) = 1$ , However,  $F_1(\infty; X) + F_2(\infty; X) \neq 1$  generally. It turns out that  $F_1(\infty; X) + F_2(\infty; X) > 1$  for some  $X$ , which can also be seen in real data (Nun Study in Section 4) analysis.

This indicates that proportional SH model doesn't hold for all competing events simultaneously.

To overcome this problem, we propose two similar but different link functions with  $g_1(F_1) = \log\{-\log(1 - F_1)\}$  and  $g_2(F_2) = \log\{-\log(1 - \frac{F_2}{1 - P_1(X)})\}$ , where  $P_1 = F_1(\infty; X)$ . Since  $F_2(\infty; X) = 1 - P_1(X)$ , it follows that  $F_1(\infty; X) + F_2(\infty; X) = 1$ .

Notice that the first link function results in a proportional SH model, which will be applied to the event of interest. An alternative way is to decompose CIF  $F_j(t; X) = P(T \leq t | \epsilon = j)P(\epsilon = j)$ , and model  $P(T \leq t | \epsilon = j)$  by using a proper distribution and  $P(\epsilon = j)$  with a binary data model (Larson and Dinse, 1985).

In the parametric modeling of CIF, the parametric form of  $\mu_j(t)$  needs to be specified. We found that  $\mu_j(t) = \log(-\log(1 - C_0 F_{0j}(t)))$ , where  $F_{0j}(t) = 1 - \exp(-(\frac{t-c_j}{b_j})^{a_j})$  is Weibull distribution function and  $C_0$  is an unknown constant ( $\leq 1$ ) fits the Nun Study data quite well.

The MLEs of parametric models are obtained by maximizing the log likelihood function:

$$\sum_{i=1}^N \left[ |l_i - 1| \sum_{j=1}^k \delta_{ji} \log\{f_j(t_i, \phi_j; X_i)\} + l_i \sum_{j=1}^k \delta_{ji} \log\{F_j(t_i, \phi_j; X_i)\} + (1 - \sum_{j=1}^k \delta_{ji}) \log\left\{1 - \sum_{j=1}^k F_j(t_i, \phi_j; X_i)\right\} \right] \quad (2.6)$$

Here for subject  $i$ ,  $X_i$  refers to a vector of covariates,  $l_i$  is an indicator for a left censored event,  $\delta_{ji}$  is an indicator for event type  $j$ ,  $\phi_j$  is an unknown parameter vector which includes Weibull parameters and regression weights for the vector  $X_i$ , and  $f_j(t, \phi_j; X_i) = \frac{dF_j(t, \phi_j; X_i)}{dt}$ . The observed information matrix can be used to evaluate the variances. The numerical computations can be done by using the nonlinear programming (NLP) procedure of SAS (see Appendix III for SAS Codes). The Hessian matrix from this computational procedure is an approximation to the Fisher Information matrix.

### 2.3 Simulation Studies

In this section we use simulations to investigate if including left censored observations improves the CIF estimation in the nonparametric case. We use additional simulations to assess the numerically computed MLEs in the parametric case. In these

simulation studies, the degree of bias and coverage probability will be assessed for each estimated parameter.

### 2.3.1 Non-parametric estimation

We generated 1000 data sets (each with sample size 1000) for 2 competing events with known failure time distribution and pre-specified censoring (both left and right levels). The first 450 observations are generated from the Weibull distribution

$$P(T \leq t | \epsilon = 1)P(\epsilon = 1) = 1 - \exp\left(-\left(\frac{t - c_1}{b_1}\right)^{a_1}\right)$$

with  $a_1 = 2.7$ ,  $b_1 = 16.5$ , and  $c_1 = 75$ . The remaining 550 observations are generated according to the Weibull distribution

$$P(T \leq t | \epsilon = 2)P(\epsilon = 2) = 1 - \exp\left(-\left(\frac{t - c_2}{b_2}\right)^{a_2}\right)$$

with  $a_2 = 3.3$ ,  $b_2 = 19$ , and  $c_2 = 74$ . Then, to specify the desired left and right censoring proportions, generate 1000 observations of  $R$  and  $L$ , from uniform distributions on intervals  $(0, R_{max})$  and  $(0, L_{max})$ , where  $R_{max}$  and  $L_{max}$  are chosen to guarantee the desired proportions of censoring. If  $R_i < T_i$ , set  $T_i = R_i$  and the failure time is right censored. If  $R_i \geq T_i$ , then if  $L_i > T_i$ , set  $T_i = L_i$  and the failure time is left censored. The choice of the Weibull parameters is motivated by the CIFs of the Nun Study data. In that data some of the event times to dementia are left censored, but no deaths are left censored. More generally, since the times to two events are different, i.e. the mean time to death and the mean time to dementia are different; the left censoring levels are likely to be different for different events. To take this fact into account, we specify different  $L_{max}$  values for event types 1 and 2 to have left censored proportion of event type 1 about twice as much as that of event type 2.

Two estimated CIFs for event 1 are computed for each dataset, one based on all the data and one based on ignoring left censored observations. Since we generate 1000 sets, the 95% point-wise confidence interval (CI) is formed by taking the range of

central 95% CIF values at each time point. As shown in Figure 2.1, 95% point-wise CI obtained by using all the data always encloses the true CIF for all left censoring levels. However, 95% point-wise CI obtained by ignoring left censored observations doesn't enclose the true CIF even in the presence of only 5% left censoring. This is due to different proportions of left censored observations was assumed for two events. Also, the width of CI obtained by using all the data is clearly smaller than those ignoring left censored data, when the left censoring level exceeds 10%, in this simulation study.

### 2.3.2 Parametric estimation

We generated 1000 data sets (each with sample size 1000) for 2 competing events with known failure time distribution, known parameters and pre-specified censoring (both left and right levels) using the probability integral transform method.

Assume the parametric form of  $F_j$  with  $g_j(F_j) = \mu_{0j}(t) + X\beta_j$ , where  $j = 1$  and  $2$ ,  $g_1(F_1) = \log\{-\log(1 - F_1)\}$  and  $g_2(F_2) = \log\{-\log(1 - \frac{F_2}{1-P_1(X)})\}$ , where  $P_1(X) = F_1(t = \infty, X)$ . Also,  $\mu_{01}(t) = C_1 F_{01}(t)$  and  $\mu_{02}(t) = F_{02}(t)$ , where  $F_{01}(t)$  and  $F_{02}(t)$  are Weibull distribution functions. Specifically,

$$F_1(t, X) = 1 - \{1 - C_1 + C_1 \exp(-\frac{t}{b_1})^{a_1}\}^{\exp(X\beta_1)}$$

and

$$F_2(t, X) = (1 - C_1)^{\exp(X\beta_1)} [1 - \{\exp(-\frac{t}{b_2})^{a_2}\}^{\exp(X\beta_2)}].$$

Note  $P_1(X) = F_1(\infty, X) = 1 - (1 - C_1)^{\exp(X\beta_1)}$ ,  $P_2(X) = F_2(\infty, X) = (1 - C_1)^{\exp(X\beta_1)}$ , and thus  $F_1(\infty, X) + F_2(\infty, X) = 1$ .

First, covariates  $X$  are simulated. Then, event indicator is created according to the marginal probabilities ( $P_1(X)$  and  $P_2(X) = 1 - P_1(X)$ ) and failure times are generated according to the following conditional probabilities using the probability

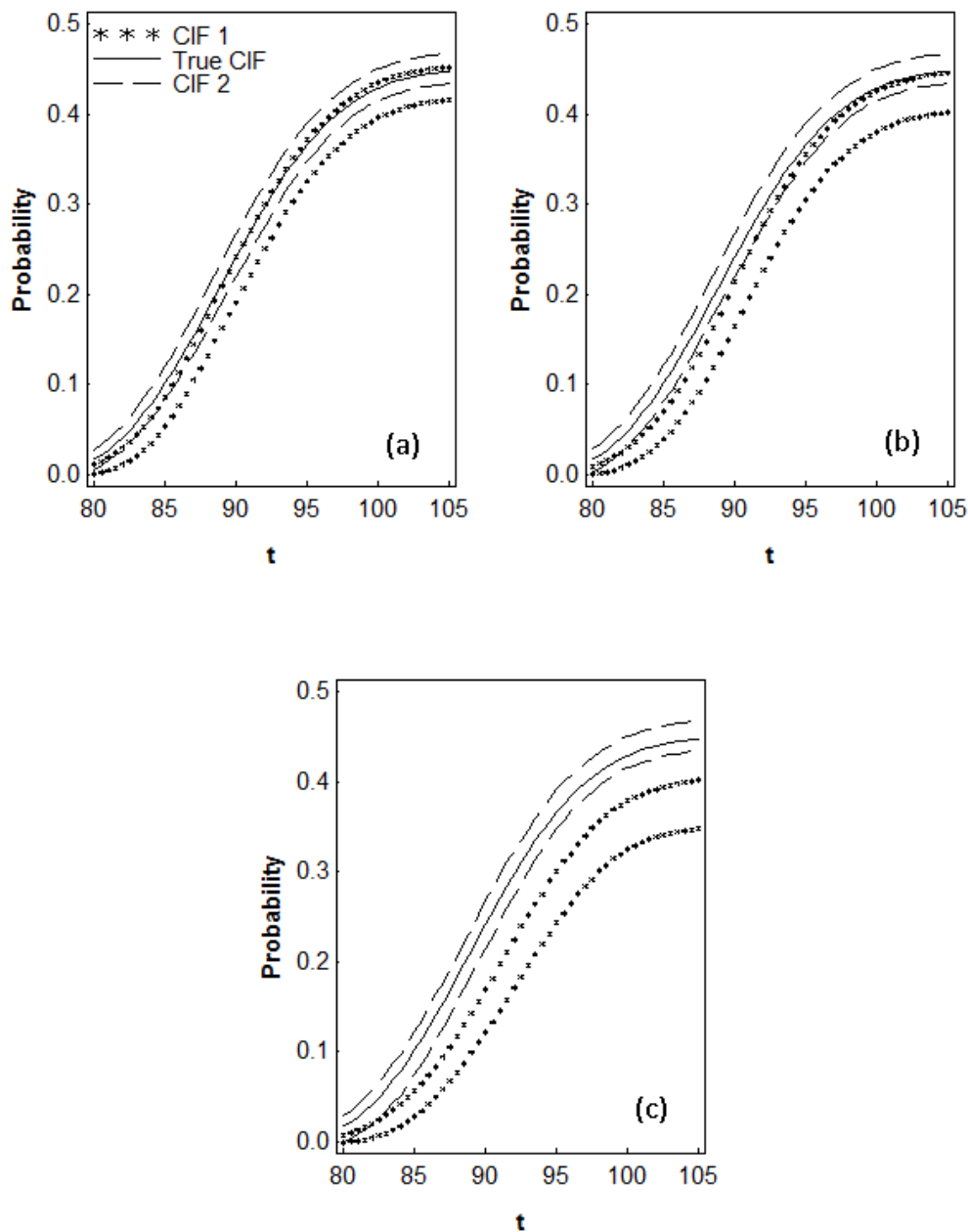


Figure 2.1: 95% Point-wise confidence interval of CIFs and true CIF for event 1 for data consisting of: (a) 20% right censoring and 5% left censoring; (b) 20% right censoring and 10% left censoring; (c) 20% right censoring and 20% left censoring. In each plot, the dash CI are obtained using all the data, the star CI are obtained by ignoring left censored observations, and the smooth curve is the true CIF used to generate the data for event 1.

integral transform method.

$$P(T \leq t | \epsilon = 1) = \frac{P(T \leq t, \epsilon = 1)}{P_1(X)} = \frac{F_1(t, X)}{1 - (1 - C_1)^{\exp(X\beta_1)}},$$

and

$$P(T \leq t | \epsilon = 2) = \frac{P(T \leq t, \epsilon = 2)}{P_2(X)} = \frac{F_2(t, X)}{(1 - C_1)^{\exp(X\beta_1)}} = 1 - \left\{ \exp\left(-\frac{t}{b_2}\right)^{a_2} \right\}^{\exp(X\beta_2)},$$

where  $a_1 = 2.6$ ,  $b_1 = 16.5$ ,  $a_2 = 3$ , and  $b_2 = 17$ .

Finally, to specify the desired left and right censoring proportions, generate 1000 observations of  $R$  and  $L$ , from uniform distributions on intervals  $(0, R_{max})$  and  $(0, L_{max})$ , where  $R_{max}$  and  $L_{max}$  are chosen to guarantee the desired proportions. If  $R < T$ , set  $T = R$  and the failure time is right censored. If  $L > T$ , set  $T = L$  and the failure time is left censored.

There are 3 covariates, i.e.  $X_i = (x_{i1}, x_{i2}, x_{i3}), i = 1$  to 1000, and  $\beta_j = (\beta_{j1}, \beta_{j2}, \beta_{j3}), j = 1$  or 2.  $x_{i1}$  takes value from Bernoulli random variable with  $p = 0.3$ ,  $x_{i2}$  takes values from discrete uniform random number  $(1, 2, 3, 4)$ , and  $x_{i3}$  are drawn from normal random variable with mean 5 and standard deviation 3. The parameters are as follows:  $\beta_{11} = 0.4, \beta_{12} = -0.2, \beta_{13} = 0.05, \beta_{21} = 0.2, \beta_{22} = 0, \beta_{23} = 0.04$ .

To address the computational accuracy, we first study the ordinary failure time data, which doesn't have any left censored observations: 5% to 35% right censoring levels are studied. The coverage probabilities for 95%, 90% and 80% CIs are summarized in Table 2.1. All the coverage probabilities match very well (within 0.01 difference) with the corresponding nominal levels, except the coverage probability of  $\beta_{12}$  (0.820) when nominal level is 0.80, under 35% right censoring. The results show that the MLEs obtained by maximization using SAS procedure NLP cover the true values very well. We also study the effect of left censored observations on the estimates. 20% left censoring and 5% to 35% right censoring levels are studied. The average bias, average standard deviation (STD), and coverage probability for 95% CI are summarized in Tables 2.2. All the coverage probabilities match very well with the



Table 2.1: Coverage probabilities for 95% confidence interval (p95), 90% confidence interval (p90), and 80% confidence interval (p80) for parameters under different right censoring levels;  $a$ 's and  $b$ 's are Weibull parameters;  $\beta$ 's are covariate effect.

Parameter	5% right censoring			10% right censoring			20% right censoring			35% right censoring		
	p95	p90	p80	p95	p90	p80	p95	p90	p80	p95	p90	p80
$C_1$	0.944	0.903	0.804	0.946	0.900	0.803	0.946	0.899	0.811	0.950	0.895	0.794
$a_1$	0.944	0.890	0.798	0.947	0.895	0.796	0.945	0.898	0.801	0.949	0.897	0.805
$b_1$	0.957	0.909	0.810	0.959	0.906	0.805	0.953	0.904	0.799	0.947	0.904	0.802
$a_2$	0.955	0.906	0.809	0.950	0.905	0.808	0.954	0.909	0.798	0.953	0.903	0.805
$b_2$	0.955	0.910	0.806	0.953	0.905	0.805	0.957	0.907	0.804	0.957	0.908	0.809
$\beta_{11}$	0.956	0.904	0.812	0.955	0.903	0.811	0.950	0.906	0.809	0.954	0.896	0.805
$\beta_{12}$	0.955	0.909	0.811	0.956	0.907	0.808	0.961	0.914	0.813	0.960	0.917	0.820
$\beta_{13}$	0.948	0.897	0.792	0.952	0.894	0.791	0.947	0.903	0.800	0.947	0.898	0.800
$\beta_{21}$	0.954	0.896	0.791	0.956	0.897	0.787	0.955	0.901	0.790	0.948	0.898	0.793
$\beta_{22}$	0.954	0.913	0.809	0.953	0.914	0.814	0.958	0.913	0.809	0.955	0.907	0.799
$\beta_{23}$	0.956	0.905	0.804	0.957	0.910	0.809	0.948	0.902	0.808	0.956	0.899	0.794

corresponding nominal level (0.95) under all right censoring levels, when estimation uses all the data. Ignoring the left censored observations, the coverage probabilities are very low for most estimates of Weibull parameters, although the coverage probabilities for estimates of covariates effects are close to 0.95. The biases of the estimates obtained by using all the data are less than those obtained by ignoring left censored data, except bias of  $\beta_{12}$ , in which case, ignoring left censored data provides a little smaller bias, though both are small. Recall the true value for  $\beta_{12}$  is a negative value (-0.2), while all the others are positive or 0. All average STD based on all the data are smaller than those estimated by ignoring left censored data, under all right censoring levels. The coverage probabilities for Weibull parameters are very low due to the fact that there is slightly different proportions of left censoring for two events. This suggests that all observations, including left censored and left truncated observations if there is any, should be used in order to obtain the correct baseline Weibull parameter estimates. However, the coverage probabilities for covariate effects are close to nominal value, even when left censored observations are excluded. This is due to the non-informative censoring, i.e., random censoring. The results show that the MLEs obtained by maximizing the likelihood based on all the data, including left censored data, are less biased, more precise, and have better coverage probabilities.

## 2.4 Application To The Nun Study

In the Nun Study, dementia and death are two competing events. The covariates include presence or absence of any copies of the APOE-4 allele (APOE4) and education. The focus of this analysis is the probability of dementia before death and the covariate effects on CIF of dementia. To look at the subgroup effect, covariate education was further divided into three groups: High, Middle, and Low, and two indicators for High and Low are used in analysis. High and Low are created according to the education level. The functions `cuminc` and `crr` in the package “`cmprsk`” from R software, which



can be downloaded from the CRAN site, were used to test equality across groups and to fit this proportional SH semi-parametric model respectively. The SAS procedure `phreg` was used to perform Cox proportional hazards modeling of each event. These methods ignore the left censoring. Results from using these methods are included here for comparison purposes and serve as starting values in obtaining MLEs using NLP procedure. We begin by examining the non-parametric MLEs of the CIFs.

#### 2.4.1 Non-parametric estimates and tests of CIFs

The non-parametric MLEs of CIFs are computed and plotted in Figure 2.2. Recall that 106 out of 606 observations are left censored observations (baseline dementias) in this study. Ignoring the left censored observations, both the incidence rate and cumulative incidence of death before dementia are higher than that of dementia, as shown in Figure 2.2(a). The estimated CIFs under both left and right censoring are obtained using 4-step iterative EM algorithm (cf. R code in Appendix II). Including the left censored data changes the relative rank of these two competing events, with incidence rate of dementia higher at younger ages (i.e. age less than 88) and lower at older ages, as shown in Figure 2.2(b). This indicates that the covariate effects might be changed after including left censored data. The data was also modeled using Weibull distributions, one for each event. These Weibull distribution functions were added into the plots with smooth curves, which fit reasonably well the non-parametric MLEs for both events, with/without left censored data. This suggests that a Weibull distribution is a good candidate as baseline distribution for parametric modeling of CIF.

The non-parametric test was performed based on Gray's test (Gray, 1988). The cumulative incidence curves of dementia and death are plotted by APOE carrier status in Figure 2.3. The tests are summarized in Table 2.3. Notice that APOE carrier status affects only the probability of a dementia before death with APOE

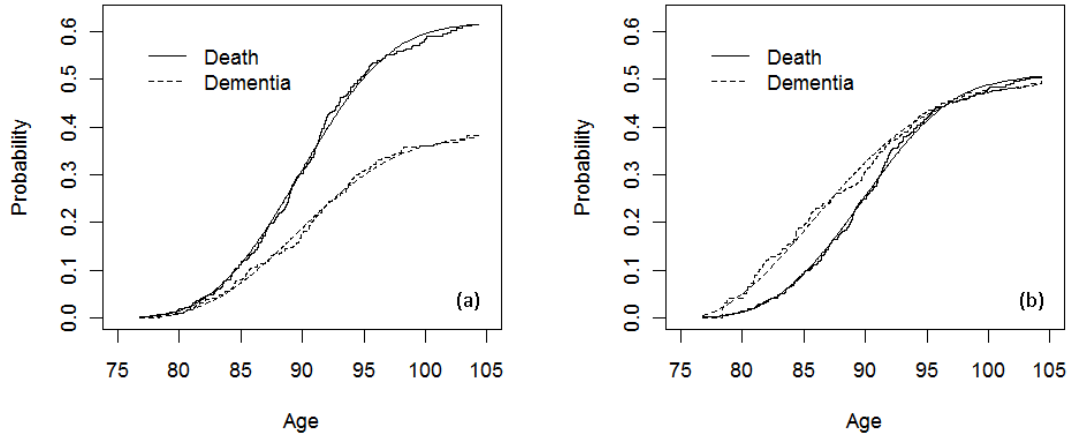


Figure 2.2: Cumulative incidence curves of dementia (solid) and death before dementia (dashed): (a) ignoring left censored observations; (b) including left censored dementia cases. The step functions are estimated CIFs using EM algorithm, while the smooth curves are fitted using Weibull distribution functions.

carrier having a significantly larger incidence for the disease.

Table 2.3: P values from Gray’s tests to compare CIFs among group variables

Event	APOE4	High	Low
Dementia	0.02	0.47	0.77
Death	0.62	0.02	0.83

### 2.4.2 Semi-parametric inferences

The `crr` function in the R package `cmprsk` fits a proportional SH semi-parametric model by maximizing the partial likelihood. The results for proportional SH model and Cox proportional hazards model are summarized in Table 2.4. Note that Cox proportional hazards model was applied to each individual event, assuming independent risks. In Cox proportional hazards model, APOE status is significant for dementia and subject with APOE carrier has significantly larger hazard (hazard ra-



Figure 2.3: Cumulative incidence function by event and by APOE4 value

tio =  $\exp(0.75) = 2.12$ ) of experiencing dementia. Higher education is significant for death and subjects with higher education have higher hazard of death. In proportional SH model, only APOE carrier status is significant for dementia, which agrees with Gray's tests; none of the covariates are significant for death. By definition, sub-distributional hazard rate is smaller than hazard rate. Because of this, the coefficients obtained in proportional SH model are normally smaller than those obtained from Cox model, assuming the coefficients are present, i.e., significant in both models. High education is significant for death in Cox model, but not in proportional SH model. This is partly due to the fact that death times are biased down by dementia times and relative higher portions of demented subjects have higher education. Usu-

ally the estimates from both methods are comparable because the proportional SH model does not take into account of the constraint we discussed before.

Table 2.4: Regression parameter estimates (standard errors) from Cox proportional hazards (Cox) and Fine and Gray’s proportional SH semi-parametric models. (number of dementia: 162; number of death before dementia: 264)

Model	Event	APOE4	High	Low
Cox	Dementia	0.75 <sup>1</sup> (0.18)	0.02 (0.17)	0.20 (0.32)
	Death	0.29 (0.16)	0.32 <sup>2</sup> (0.13)	0.21 (0.26)
F - G	Dementia	0.49 <sup>2</sup> (0.18)	-0.21 (0.16)	0.01 (0.33)
	Death	-0.08 (0.17)	0.22 (0.13)	0.05 (0.25)

<sup>1</sup>P-value less than 0.001

<sup>2</sup>P-value less than 0.01

### 2.4.3 Parametric modeling of CIF

However, neither of these two semi-parametric models can accommodate left censored data. The left censoring is well handled in the parametric models with  $F_{j0}(t) = 1 - \exp(-\frac{t-c_j}{b_j})^{a_j}$ . Proportional SH model subject to the constraint is of another interest. Being aware that the dementia times are relatively smaller than death times since the death times of dementia subjects are obviously larger than their dementia times. Besides this dementia is our focus event. With these considerations, we always assume proportional SH for dementia after taking into account of the constraint that overall probability of any event is 1. The fit of three different parametric models are summarized in Table 2.5. In model (1), which assumes proportional SH for both events (Jeong and Fine’s parametric model), APOE status is highly significant for the SH of dementia and higher education is significant for the SH of death. All the others are not significant. As the same problem which appears in Fine’s semi-parametric proportional SH model, probability of experiencing any event might be greater than 1 in this model. According to the estimated parameters and assuming a subject with APOE carrier and low education,  $F_{dementia}(\infty; Apoe4 = 1, low =$

Table 2.5: Regression parameter estimates (standard errors) from parametric models for cumulative incidence of dementia and death; Model (1) ignores left censoring and assumes proportional SH for both dementia and death, resulting in probability of experiencing any event being greater than 1; Model (2) ignores left censoring and assumes proportional SH for dementia; Model (3) includes left censored data and assumes proportional SH for dementia.

Model	Event	b	a	c	APOE4	High	Low	Event Pr.
(1)	Dementia	17.81 (2.01)	2.71 (0.44)	75.63 (1.79)	0.85 <sup>3</sup> (0.20)	-0.30 (0.17)	0.48 (0.31)	0.40 (0.04)
	Death	18.41 (3.16)	3.16 (0.37)	74.62 (1.42)	0.04 (0.16)	0.36 <sup>2</sup> (0.13)	0.30 (0.26)	0.60 (0.04)
(2)	Dementia	16.35 (1.78)	2.67 (0.40)	75.89 (1.57)	0.40 <sup>1</sup> (0.18)	-0.13 (0.16)	0.02 (0.32)	0.38 (0.04)
	Death	19.00 (1.72)	3.34 (0.40)	74.32 (1.53)	0.45 <sup>2</sup> (0.17)	0.38 <sup>2</sup> (0.14)	0.09 (0.27)	
(3)	Dementia	14.57 (2.79)	2.13 (0.45)	74.96 (2.53)	0.58 <sup>3</sup> (0.13)	-0.43 <sup>2</sup> (0.14)	0.49 <sup>2</sup> (0.19)	0.49 (0.03)
	Death	18.98 (1.71)	3.33 (0.40)	74.34 (1.52)	0.47 <sup>2</sup> (0.17)	0.35 <sup>2</sup> (0.14)	0.12 (0.27)	

<sup>1</sup>P-value less than 0.05

<sup>2</sup>P-value less than 0.01

<sup>3</sup>P-value less than 0.0001



1) =  $1 - [1 - 0.402]^{\exp(0.85*1+0.48*1)} = 0.857$  and  $F_{death}(\infty; Apoe4 = 1, low = 1) = 1 - [1 - 0.598]^{\exp(0.044*1+0.304*1)} = 0.725$ . These lead to

$$P(T < \infty) = 0.857 + 0.725 = 1.582 > 1,$$

which makes the proportional SH model a problematic model. The modified models assume proportional SH for only the event of interest. When assuming proportional SH for dementia and ignoring the left censored data (model (2)), APOE status is still significant, which agrees with Fine's semi-parametric model on dementia, but at a much reduced risk. Note that in models (2), only the event dementia for which proportional SH was assumed is comparable to Fine's semi-parametric proportional SH model. Thus the coefficients for death do not have the same meaning as those in proportional SH model. When assuming proportional SH for dementia only and including the left censored data (model (3)), besides APOE status, both high and low educations are significant. Considering three education groups, the higher the education the less the risk for dementia. The effect of APOE becomes much larger. Again, due to the consideration of the constraint, the coefficients for death are not meaningful in model (3). Standard error estimates for all covariates from model (3) are smaller comparing to those from model (1), model (2), Fine's semi-parametric model, or Cox model, which did not include left censored data. Therefore, including left censored data into the analysis, parameter estimates are more precise.

## 2.5 Discussion

In this Chapter, we proposed a strategy for analyzing competing risks data in the presence of left censoring. Parametric modeling on CIF was first proposed by Jeong and Fine (2006) by specifying complementary log-log link function for CIF, which is linked to Fine's semi-parametric proportional SH model and recognized also as a proportional SH model. However, allowing proportional SH for all competing events

leads to a problem, i.e.  $P(T < \infty) > 1$ . Our modified parametric model assumes proportional SH for the event of interest and solves the above problem. The disadvantage of such modeling is that the interpretation for the other events is not straightforward. Both simulation studies, with/without covariates, show that including left censored data improves the estimates, resulting in less bias, better coverage probability and more preciseness. The iterative EM algorithm can be used to obtain the non-parametric MLEs of CIFs when the competing risks data is also subject to left censoring. Neither Fine’s semi-parametric model nor Cox proportional hazards model can accommodate left censored data. This is not the case for parametric models since estimates are obtained by maximizing a likelihood, which is always defined even in the presence of left censoring.

We then applied the above models to Nun’s data in the presence of left censoring.

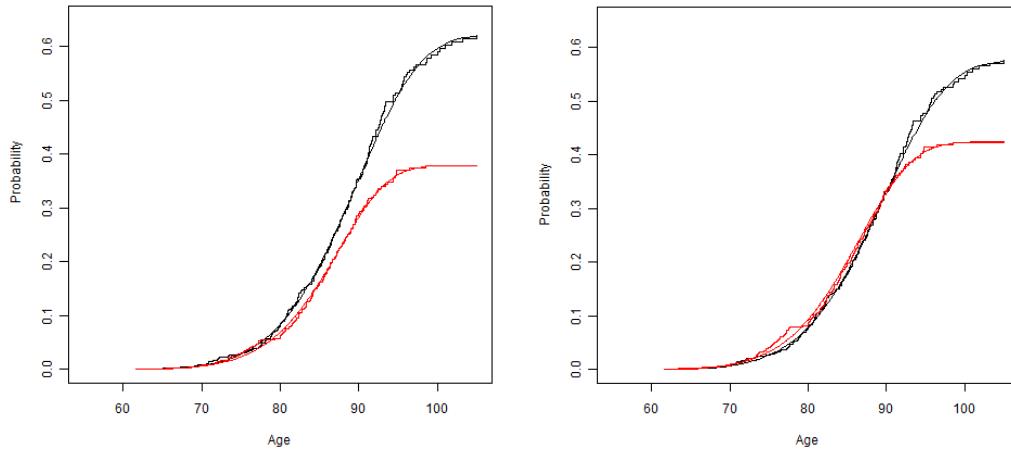


Figure 2.4: Cumulative incidence curves of death (black) and MCI (red). The smooth curves are fitted curves using Weibull distribution function. (the top graph excludes left and interval censored data, while the bottom graph includes those. )

The baseline sub-distribution was modeled using the product of an unknown constant and a Weibull distribution, which accommodates improper distributions of CIFs and gives the marginal probability of event of interest as a by-product. We found both APOE carrier status and education are significant for incidence of dementia, with

education being protective.

Parametric analysis handles left censored data and enables estimation of the long-term proportion of individuals experiencing a particular event type using a simplified parametric distribution. Weibull distribution was used to parameterize the baseline distribution. It is worth noting that Weibull distribution could fit CIF with or without a plateau. Such flexibility enables it to be applied for other types of competing risks data. The figure 2.4 shows how well Weibull distributions fit to the baseline CIFs of BRAiNS data with MCI and death as competing events. The goodness-of-fit of baseline distribution was examined by simple plots. It will be of great interest to propose a more specific goodness-of-fit test.

## Chapter 3 Multi-state Modeling

Multi-state Markov model is often used to analyze the evolution of cognitive states. However, the Markovian assumption is violated in BRAiNS data since some transitions are likely time and duration dependent while the others are time homogeneous, leading to a mixture of semi-Markov and Markov processes. We adopted Larson and Dinse (1985) competing risks mixture model and broke the joint probability of observing a sequence of same state until transition into two components: one for the overall transition probability and the other for survival probability. These two quantities were simultaneously modeled for semi-Markov process. Such modeling leads to different interpretations in BRAiNS study, i.e., family history, APOE4, and sex by head injury interaction are significant factors for transition intensities in traditional Markov model. While in our semi-Markov model, these factors are significant in predicting the overall transition probabilities, but none of these factors are significant for duration time distribution.

### 3.1 Introduction

The typical research of Alzheimer's disease, a chronic disease, includes a series of cognitive assessments of subjects over a period of years which are observed at certain time points, for example, every 2 years. The cognitive assessment evaluates memory, language, ability to recognize and other cognitive abilities. Such longitudinal studies record the progression of healthy individuals to chronic diseases, such as cognitive impairment or dementia. Continuous-time, multi-state process are often used to describe the history of disease evolution. In this chapter we consider not only the absorbing states, but also transient and initial states. Panel data are realizations of a continuous-time process at arbitrary times. One particular example is homogeneous

Markov multi-state model which characterizes transition intensities between states by a set of time-constant.

However, the state structure is not unique and even the state definitions are different across different centers. Choosing the the appropriate state structure will simplify the calculation and provide useful information. In the Biologically Resilient Adults in Neurological Studies (BRAiNS) at the University of Kentucky’s Alzheimer’s Disease Center (UK ADC), a longitudinal cohort of 1,030 individuals established in 1989 with ongoing recruitment. Participants consent to extensive annual cognitive and clinical examinations as well as brain donation upon death. Exclusion criteria include age less than 60 years, active infectious diseases, neurological disorders, psychiatric disorders, disabling medical disorders, and dementing illness.

Annual cognitive assessments taken on a cohort of initially cognitively intact subjects participating in the BRAiNS project are used to classify subjects into one of three states: normal healthy, Clinical MCI, or dementia. Between assessments subjects may die or become demented, and these states are treated as completely absorbing competing states. Clinical MCI is treated as a quasi-absorbing state, as subjects do not move backward to normal healthy, but they may become demented or die. Classification into clinical MCI results from a diagnosis of MCI, which is determined according to the consensus guidelines on MCI developed by the Second International Working Group on MCI (Winblad, et al, 2004). A dementia classification also results from a clinical consensus diagnosis of dementia (most often AD). Schmitt, et al (2001) and Abner, et al (2011) described more details of the BRAiNS project. In this study, the following four states are used to describe the disease process through the graphical model, as shown in Figure 3.1.

1. Healthy normal.
2. Clinical mild cognitive impairment(MCI): Diagnosis of clinical MCI is based on a consensus team review by the examining physicians, neuropsychologists, and the

clinical research assistant who administered the testing protocol (Jicha, et al, 2008).

3. Dementia: which is based on the criteria of the Joint Working Group of the National Institute of the Neurologic and Communication Disorders and Stroke-AD and Related Disorders (NINCDS-ARDA) (McKhann, et al, 1984).

4. Death.

Health is the initial state for every subject, MCI is the transient state, death

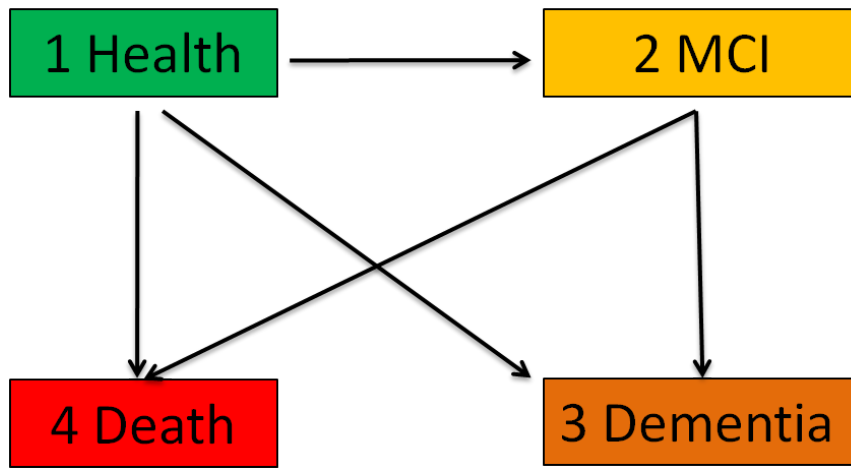


Figure 3.1: Four-state graphical model of BRAiNS study

and dementia are absorbing states. None of these events are recurrent. As introduced in Chapter 1, there is considerable interest to model and inference transition probabilities and state occupation probabilities, since probabilities are simple to interpret. Such models are usually done by assuming a Markov process and there are corresponding transition intensity models. However, transition intensities may vary between individuals. They could be time-dependent which gives an inhomogeneous Markov model or non-Markov model, but there is no easy solution to probabilities. Modeling transition intensity is more flexible since the multi-state models based on intensities can borrow the models known from survival analysis. Our data can be fully described by transition intensities. The most simple model assumes a time-homogeneous Markov process, where each transition is characterized by a time-

constant transition intensity. Such a simple model is not appropriate because clearly, the hazard rate (or transition intensity) of death from health is age dependent. Thus a model with time dependent transition intensities is required.

Models fully based on transition intensities have the underlying assumption that all subjects have the same hazard rate of each event, if no covariate is considered. This might not be the case in the BRAiNS data, since we think there is a group of subjects with particular genotype who are likely to convert to MCI from health and another group who will never convert to MCI. Distinguishing these two will help us better understand the true hazard rates for those who will develop MCI, i.e. the time to MCI. Therefore we propose a mixture model with logistic model for probabilities and survival model for time to events, leading to a semi-Markov process.

In this chapter, we apply Bayesian mixture models to longitudinal failure time data with the consideration of following issues: (1) missing data, subjects who are alive and healthy have unknown event type; (2) interval censoring, because of the yearly visits, event times are subject to interval censoring; (3) random effect, subject-specific random effect will account for unobserved heterogeneity. Notice that the likelihood function for multi-state models are usually constructed based on transition probabilities. Therefore, direct modeling of transition probabilities avoids compute transition probabilities from transition intensities. The likelihood for the multi-state model is constructed through a multinomial distribution. The estimation are obtained through Bayesian approach based on Markov chain Monte Carlo(MCMC) simulation technique using OpenBUGS. As an application, we show the covariate effects on transition intensities, overall transition probabilities, and duration times in the BRAiNS data. We then apply our models to BRAiNS data and compare with the results obtained from traditional intensity-only model.

The structure of this chapter is as follows: Section 2 describes the statistical models specification and likelihood construction. In Section 3 we explain the Bayesian esti-

mation method and model selection criteria. In Section 4, we apply these methods to the BRAiNS data, where we first use a homogeneous Markov process without covariates and compare it to a piecewise homogeneous Markov model. Then parametric baseline transition intensities, the effects of certain covariates and random effect are introduced into the model. Finally a discussion is given in Section 5.

### 3.2 Model Specification

The multi-state process is fully described by transition intensities, the matrix of which  $Q(t)$  is expressed as follows:

$$Q(t) = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) & \alpha_{13}(t) & \alpha_{14}(t) \\ 0 & \alpha_{22}(t) & \alpha_{23}(t) & \alpha_{24}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

where  $\alpha_{hj}(t)$  is defined by

$$\alpha_{hj}(t) = \lim_{\Delta t \rightarrow 0} \frac{P_{hj}(t, t + \Delta t)}{\Delta t} = \lim_{\Delta t \rightarrow 0} \frac{pr\{X(t + \Delta t) = j | X(t) = h\}}{\Delta t}$$

for  $h, j \in \{1, 2, 3, 4\}$ . The corresponding transition probability matrix is

$$P(t) = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} P_{11}(t) & P_{12}(t) & P_{13}(t) & P_{14}(t) \\ 0 & P_{22}(t) & P_{23}(t) & P_{24}(t) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

Within a given time interval  $t$ , let  $n_{jj'}$  represent the observed number of subjects who moved from state  $j$  to  $j'$ ,  $P_{jj'}$  represent the probability of transition from state  $j$  to



$j'$ , also the trial number  $n_j = 1$  and number of states  $k = 4$ . Then

$$(n_{j_1}, \dots, n_{j_4}) \sim \text{Multinomial}(P_{j_1}, \dots, P_{j_4}; 1).$$

The overall likelihood is:

$$L = \prod_{i=1}^N \prod_{l=1}^{v_i} \left[ \prod_{j=1}^4 \prod_{j'=1}^4 P_{jj'}(t_{i,l})^{n_{jj'}} \right]$$

The remaining work is to specify a model for each transition probability. Depending on the assumption of Markov or semi-Markov process, the transition probability can be expressed differently. The following table shows observations from one subject. The joint probability of observing such a sequence of states for the subject can be

Table 3.1: A simple example of record of states: 1(health), 2(MCI), 3(dementia)

Visit #	1	2	3	4	5	6	7	8
State	1	1	1	1	2	2	2	3
Age	68	70	72	74	76	78	80	82

expressed in a sequence of conditional probabilities:

$$\begin{aligned} & P(s_8 = 3, t_8 = 82, s_7 = 2, t_7 = 80, \dots, s_1 = 1, t_1 = 68) \\ = & P(s_8 = 3, t_8 = 82 | s_7 = 2, t_7 = 80, s_6 = 2, t_6 = 78, \dots, s_1 = 1, t_1 = 68) \\ & * P(s_7 = 2, t_7 = 80 | s_6 = 2, t_6 = 78, s_4 = 1, \dots, s_1 = 1, t_1 = 68) \\ & * \dots * P(s_2 = 1, t_2 = 70 | s_1 = 1, t_1 = 68) * P(s_1 = 1, t_1 = 68) \end{aligned}$$

Under Markov assumption, the transition probability depends on only the current state, thus, the above joint probability becomes

$$\begin{aligned} & P(s_8 = 3, t_8 = 82 | s_7 = 2, t_7 = 80) * P(s_7 = 2, t_7 = 80 | s_6 = 2, t_6 = 78) \\ & * \dots * P(s_2 = 1, t_2 = 70 | s_1 = 1, t_1 = 68) * P(s_1 = 1, t_1 = 68) \end{aligned}$$

The above joint probability is constructed under the Markovian assumption. However, it is very likely that the transition probabilities depend on the duration time of the

previous state. To relax the assumption but not lose any information is to consider only those times when state changes, resulting in a data consisting of time to events (change of states). In our case, the data consists of two sets of competing risks data. MCI, dementia, and death are competing risks for healthy normals, dementia and death are competing risks for subjects who have converted to MCI. Similar to the models discussed in Chapter 2, assume independent risks and apply survival model for each event, or apply our proposed competing risks model on joint probability without assuming such independence. But clearly, MCI and death form a semi-competing risks data. Therefore, we adopt Larson and Dinse's model and divide the joint density into the product of marginal density for event probability and conditional density for event times. The transition probability can be broken into two components: one for overall transition probability and the other for survival probability conditional on the specified transition, leading to a semi-Markov process. Then, the joint probability can be written as

$$\begin{aligned}
& P(T_{23} \leq 82 | T_{23} \geq 80) * P(T_{23} \geq 80 | T_{23} \geq 78) * P(T_{23} \geq 78 | T_{23} \geq 76) \\
& * p(2 \rightarrow 3) \\
& * P(T_{12} \leq 76 | T_{12} \geq 74) * \dots * P(T_{12} \geq 70 | T_{12} \geq 68) \\
& * p(1 \rightarrow 2) \\
= & P(80 \leq T_{23} \leq 82) * p(2 \rightarrow 3) * P(74 \leq T_{12} \leq 76) * p(1 \rightarrow 2)
\end{aligned}$$

The survival model works for each event time and multinomial logistic model works for the overall transition probability. The exact transition times sometimes are known, for example, the time of death is known in the observational studies of chronic diseases. If this is the case, the corresponding transition probability is replaced by transition intensity. The transition intensity by definition is very similar to the hazard rate in a survival model. When exact transition times are known, the likelihood constructed based on transition matrix is the same as that constructed based

on counting process theory assuming independent risks (Kneib, 2008).

These two methods construct likelihoods differently, especially when covariate effects are taken into account. We think the second method is more appropriate for the BRAiNS study with the following considerations: 1. The transitions might be duration dependent; 2. Covariate effects on probability of being demented and on time to dementia are different.

Both methods require specifying the model for transition intensity or hazard rate. Let  $h = 1, \dots, H$  index the type of transition and  $i = 1, \dots, N$  index the individual subjects. We adopt a proportional hazards specification. The model specification for the conditional hazards on event is given by a generalized Cox proportional hazards models. Another regression model for survival data that extends to multi-state models is the non-parametric (or parametric) additive model (Buckley, 1984).  $Z_i$  is a vector that contains explanatory variables (covariates);  $\alpha^{(h)}$  is the hazard rate for transition type  $h$ ;  $t$  is the time from the beginning of last event;  $\beta$  is a vector of unknown parameters representing fixed effects. For the sake of simplicity, we will drop the transition index and  $\alpha(t|Z_i)$  represents the hazard rate for subject  $i$ . In order to investigate the baseline hazard rates, we examine time homogeneous, piecewise constant, and parametric baseline hazard rate. Furthermore, we add the frailty term to account for the possible heterogeneity.

***Time homogeneous hazard rate***

$$\alpha(t|Z_i) = \alpha_0 \exp(\beta^T Z_i)$$

where  $\alpha_0$  is an unknown constant.

***Piecewise constant hazard rate***

$$\alpha(t|Z_i) = \alpha_0(t) \exp(\beta^T Z_i)$$

where  $\alpha_0(t)$  is a step function of time  $t$ .

***Parametric hazard rate***

$$\alpha(t|Z_i) = \alpha_0(t)\exp(\beta^T Z_i)$$

where  $\alpha_0(t) = a * (\frac{t-c}{b})^{a-1}$ , where  $a > 1$ ,  $b > 0$ , and  $c > 0$ . This is obtained from assuming a Weibull distribution function for survival probability  $F(t)$ . Note that  $H_0(t)$  is cumulative baseline hazard function.

$$\begin{aligned} F(t) &= 1 - \exp(-(\frac{t-c}{b})^a) \\ S(t) &= \exp(-(\frac{t-c}{b})^a) \\ H(t) &= (\frac{t-c}{b})^a \\ \alpha_0(t) &= \frac{a}{b} * (\frac{t-c}{b})^{a-1}. \end{aligned}$$

***Parametric hazard rate with shared random effect***

$$\alpha(t|Z_i) = \alpha_0(t)\exp(\beta^T Z_i + b_i)$$

where  $\alpha_0(t) = a * (\frac{t-c}{b})^{a-1}$  and  $b_i$  is the subject-specific random effect.

These specifications encompass, for example, the homogeneous Markov chain model that is frequently used. These models can be extended to model duration dependence transitions, resulting in a semi-Markov model.

***Multinomial logistic model for the overall transition probabilities***

$$p(j \rightarrow j'|Z_i^*) = \frac{\exp(\gamma_{jj'}^T Z_i^*)}{\sum_{j' \neq j} \exp(\gamma_{jj'}^T Z_i^*)},$$

where  $j' \neq j$ ,  $\gamma$  is a vector of unknown parameters representing fixed effects, including intercept, and  $Z_i^*$  is a vector of selected covariates, which can be different from the covariate vector  $Z_i$  used in hazard modeling.

**3.3 Estimation and Model Selction**

Usually the death times are observed exactly, however, the times to MCI and dementia are interval censored. The interval censored times to MCI cause the the time from

MCI to death to be interval censored as well. For the subjects who are still alive and MCI free currently or the subjects who have experienced MCI but are alive and dementia free, the type of event is unknown and the time is right censored. Besides this missing information, the subject-specific random effect will also introduce a number of parameters. With the consideration of event type imputation for those unknown event types, the MCMC method is used to obtain the parameter estimates (Chen, Shao, and Ibrahim, 2000).

OpenBUGS is used for generating samples. The post-process of the simulation results is tedious since the user needs to specify all the parameters to be monitored in each run, one by one. It is thus quite useful to call OpenBUGS from R, which is used to input data and initial values, specify the parameters to be monitored (only once), and perform other post process. The R package “BRugs” (R development Core Team, 2007) does the work, together with the other package “arm”. (See Appendix V for OpenBUGS program codes.)

The common approach of obtaining “noninformative” prior is assuming the unknown parameters are independent of each other and the joint prior can be specified as the product of individual priors. We assign weakly informative priors to make sure that estimations are driven by the observed data. In particular, univariate normal priors with large variation are assigned for covariate effects, uniform priors are assigned for Weibull parameters, gamma priors are assigned for piecewise hazard rate and random effect.

Deviance is defined as  $D(\theta) = -2\log(p(y|\theta)) + C$ , where  $y$  are the data,  $\theta$  are the unknown parameters of the model and  $p(y|\theta)$  is the likelihood function. Note  $C$  is a constant that will be canceled out in model comparisons. Deviance Information Criterion(DIC)(Spiegelhalter, et al., 2002), which is intended as a generalization of Akaike’s Information Criterion (AIC), is a Bayesian method for model comparison that OpenBUGS can easily calculate from the samples generated by a MCMC simu-

lation for many models. DIC is given by

$$DIC = \bar{D} + pD,$$

where  $\bar{D}$  is the posterior mean of the deviance,  $pD$  is 'the effective number of parameters' with  $pD = \bar{D} - \hat{D}$ .  $\hat{D}$  is a point estimate of the deviance obtained by substituting in the posterior means  $\bar{\theta}$  thus  $\hat{D} = -2 * \log(p(y|\bar{\theta}))$ . For simple models with little prior information,  $pD$  should be approximately the number of parameters. The model with the smallest DIC is estimated to be the model that would best fit the dataset that is currently observed.

### 3.4 Application to the BRAiNS Study

543 Subjects with non-missing covariates of interest in the BRAiNS study are included in this analysis. Besides cognitive state, the times to MCI, dementia, or death are our interested response variables too. The covariates of interest include sex, presence or absence of any copies of the APOE-4 allele(APOE4), presence or absence of family history of dementing illness among first degree relatives (family history), and presence or absence of history of hypertension or head injury at entry (head injury). The interaction between history of head injury and sex was also tested, as a head injury of men and women are possibly different.

#### 3.4.1 Markov model using “msm” package

The main focus of this study is to identify risk factors. But we will start by looking at time-homogeneous and piecewise Markov models without covariates. Comparing these two models will tell us whether assuming time-homogeneous transition intensities is realistic. The Markov modeling of panel data can be performed by R package

“msm” (Jackson, 2011). A sample R code of fitting Markov model using “msm” is attached in Appendix IV. Table 3.2 shows the frequency counts for each transition. Based on this table, we specify the possible transitions by assigning non-zero initials.

Table 3.2: Transition numbers for BRAiNS data

Prior state	Current state			
	Healthy	MCI	Dementia	Death
Healthy	4000	97	51	194
MCI	0	100	31	15

The resulting homogeneous transition intensity matrix is shown in Table 3.3. Healthy normals have about the same risks of converting to MCI (hazard rate = 0.034) and death before MCI (hazard rate = 0.039). But individuals who already converted to MCI are more likely to get demented before death. The hazard rate of converting to dementia is about 3 times as that of death.

Table 3.3: Transition intensity matrix (enclosed in parentheses are 95% confidence interval). Not-observed transitions have zero intensities

	1	2	3	4
1	-0.07(-0.08,-0.06)	0.03(0.03,0.04)	0	0.04(0.03,0.05)
2	0	-0.39(-0.48,-0.32)	0.30(0.24,0.37)	0.09(0.06,0.16)
3	0	0	0	0
4	0	0	0	0

As the transition intensity matrix ( $Q$ ) is constant over time in the homogeneous process, the transition probability matrix at time  $t$  ( $P(t)$ ) can be easily obtained by solving the equation:  $P(t) = \exp(tQ)$ . For example  $P(5)$  is computed and shown in Table 3.4. Five years after first visit, the probability of being MCI is 5.9%, dementia 6.3%, and death 18%. the probability of converting to dementia after being in MCI is 65%.

Instead of assuming time homogeneous transition intensities, piecewise constant in-

Table 3.4: Transition probability matrix 5 years after first visit

	1	2	3	4
1	0.70(0.67,0.72)	0.059(0.048,0.071)	0.063(0.051,0.077)	0.18(0.16,0.21)
2	0	0.14(0.093,0.20)	0.65(0.55,0.73)	0.21(0.13,0.31)
3	0	0	1(1,1)	0
4	0	0	0	1(1,1)

tensities were fit to the data as well. It turns out that model fit by piecewise constant intensities results in much larger likelihood ( $-2 \log LR = 91.1$  with 8 degree of freedom,  $p\text{-value} < 0.00001$ , compared with time homogeneous intensities). This indicates that transition intensities are significantly time dependent.

The covariate effects were then investigated. For easy understanding, time homogeneous Markov model was used and again, the analysis was done by using “msm” package. Age at baseline was also included as a covariate. The results for hazard ratios are summarized in Table 3.5

Note that there are quite a few transitions from MCI to dementia or death and the variation for covariate effects is quite large. Age at baseline is significant for all transitions except MCI to death, the hazard increases as the baseline age increases. Family history is only significant for the transition from MCI to dementia, subjects with family history have higher hazard (ratio=1.7). APOE4 is only significant for the transition from healthy to MCI (hazard ratio=1.7). Females have high hazard of dementia after experiencing MCI (hazard ratio=1.8). Also, females who had head injury are at higher risk of MCI.

### 3.4.2 Time homogeneous hazard rates

Considering that the transitions from health to dementia experiences MCI before converting to dementia, the transitions from health to MCI and dementia are combined and the transition times to dementia are considered as left censored times for transition to MCI. The models are set-up by assuming logistic models for event prob-



Table 3.5: The hazard ratio of each factor for each transition, resulting from time-homogeneous Markov model fit by R package “msm”

Covariate	Transition type	Hazard ratio	95% Lower	95% upper
Age at baseline	Health - MCI	1.049	1.025	1.073
	Health - Death	1.103	1.080	1.126
	MCI - Dementia	1.053	1.016	1.091
	MCI - Death	0.937	0.856	1.025
Family history	Health - MCI	1.091	0.762	1.561
	Health - Death	0.719	0.495	1.044
	MCI - Dementia	1.669	1.017	2.740
	MCI - Death	0.905	0.279	2.934
APOE4	Health - MCI	1.730	1.225	2.443
	Health - Death	0.819	0.555	1.208
	MCI - Dementia	0.808	0.491	1.330
	MCI - Death	1.173	0.362	3.801
Sex	Health - MCI	0.776	0.530	1.135
	Health - Death	0.769	0.534	1.107
	MCI - Dementia	1.771	1.005	3.119
	MCI - Death	0.588	0.177	1.960
Head injury	Health - MCI	0.542	0.270	1.086
	Health - Death	1.279	0.801	2.044
	MCI - Dementia	0.999	0.331	3.014
	MCI - Death	0.516	0.0652	4.083
Sex*	Health - MCI	3.925	1.631	9.445
Head injury	Health - Death	0.699	0.282	1.734
	MCI - Dementia	1.137	0.312	4.150
	MCI - Death	12.925	0.965	173

abilities and survival models for event times. Note that treating time to death from healthy as right censored observations when time to MCI is our primary interest will under-estimate the hazard of converting to MCI. i.e. over-estimate the time to MCI, for those who will experience MCI before death. On the other hand, treating time to MCI as right censored observations when time to death (from healthy) under-estimate the survival times. Thus it is important to use time to MCI survival model for only those who will convert to MCI and time to death survival model for only those who will die without converting to MCI.

Four binary covariates, i.e., Family history, APOE4, Sex, Head injury, and interaction between Sex and Head injury are included in both logistic models of probabilities and survival models. The baseline hazards for survival models are unknown constant. For  $i$ th subject,  $D_{1i}$  and  $D_{2i}$  indicate the event types.

$$D_{1i} = \begin{cases} 1 & , \text{ MCI (from Health)} \\ 2 & , \text{ Death (from Health)} \\ \text{NA} & , \text{ Alive and Healthy} \end{cases}$$

and

$$D_{2i} = \begin{cases} 1 & , \text{ Dementia (from MCI)} \\ 2 & , \text{ Death (from MCI)} \\ \text{NA} & , \text{ otherwise} \end{cases}$$

The logistic model for probability of converting to MCI from healthy state is given by

$$\begin{aligned} \text{logit}(P(D_{1i} = 1)) &= \text{int}[1] + \text{ap}[5] * \text{apoe}4_i + \text{fa}[5] * \text{famhx}_i + \text{se}[5] * \text{sex}_i \\ &\quad + \text{hd}[5] * \text{headinj}_i + \text{sh}[5] * \text{sex}_i * \text{headinj}_i \end{aligned} \quad (3.1)$$

The logistic model for probability of converting to dementia from MCI is given by

$$\begin{aligned} \text{logit}(P(D_{2i} = 1)) &= \text{int}[2] + \text{ap}[6] * \text{apoe}4_i + \text{fa}[6] * \text{famhx}_i + \text{se}[6] * \text{sex}_i \\ &\quad + \text{hd}[6] * \text{headinj}_i + \text{sh}[6] * \text{sex}_i * \text{headinj}_i \end{aligned} \quad (3.2)$$

Note  $P(D_{1i} = 2) = 1 - P(D_{1i} = 1)$  and  $P(D_{2i} = 2) = 1 - P(D_{2i} = 1)$ .

The survival model for time to events are constructed based on hazard function, as shown in the following equations. The likelihood construction takes into account of right censoring and interval censoring.

$$\begin{aligned}
H_{k0,i} &= \alpha_{k0,i} * t_{k,i} \\
H_{k0,i} &= \alpha_{k0,i} * t_{k,i} \\
H_{k,i} &= H_{k0,i} * \exp(ap[k] * apoe4_i + fa[k] * famhx_i + se[k] * sex_i \\
&\quad + hd[k] * headinj_i + sh[k] * sex_i * headinj_i) \\
F_{k,i} &= 1 - \exp(-H_{k,i}) \\
f_{k,i} &= \exp(-H_{k,i}) * \exp(ap[k] * apoe4_i + fa[k] * famhx_i + se[k] * sex_i \\
&\quad + hd[k] * headinj_i + sh[k] * sex_i * headinj_i) \\
\log l_{k,i} &= (1 - d_{k,i}) * H_{k,i}^U + d_{k,i} * \log((1 - I_{k,i}) * f_{k,i}^U + I_{k,i} * (F_{k,i}^U - F_{k,i}^L))
\end{aligned}$$

Where  $k = 1, 2, 3, 4$ .  $t_{k,i}(t_{k,i}^L, t_{k,i}^U)$  is the time to event  $k$ , which is subject to interval censoring. The overall likelihood for survival models is the sum of likelihood for experienced event types. For example, a subject whose cognitive process follows “Health to Health to MCI to MCI to Dementia” experiences event type 1 (healthy to MCI) and 3 (MCI to dementia). If a subject doesn’t experience any event, for example, is healthy and alive at the end of study, then this subject could experience event type 1 or 2 (healthy to death), depending on the imputed event type which is done through data augmentation in MCMC data updating. Data augmentation for response variables is automatically performed in OpenBUGS.

Table 3.6 shows the parameter estimates and their SDs and 95% confidence intervals for logistic models and survival models. APOE4 and family history are significant for predicting the probability of converting to MCI from normal healthy, with APOE-4 allele present having higher hazard rate (hazard ratio =  $\exp(0.74)$ =2.10) and family history present having higher hazard rate (hazard ratio =  $\exp(0.53)$ =1.70). However, none of the covariates included are significant for any survival models. The SDs are large in survival models for time to death or dementia from MCI, this is due to the very limited cases of transitions from MCI to death or dementia observed in the study

Table 3.6: Parameter estimates under time homogeneous hazard rates

Parameter	Transition type	Mean	SD	95% Lower	95% upper
Logistic modeling of probability of MCI					
Family history		0.53	0.27	0.01	1.05
APOE4		0.74	0.27	0.22	1.27
Sex		0.08	0.29	-0.49	0.64
Head injury		-0.8	0.45	-1.69	0.06
Sex*Head injury		1.03	0.58	-0.1	2.2
Logistic modeling of probability of Dementia					
Family history		0.67	0.82	-0.94	2.30
APOE4		-1.05	0.72	-2.54	0.31
Sex		0.07	0.79	-1.47	1.64
Head injury		-0.79	1.85	-4.3	3.0
Sex*Head injury		1.15	2.11	-2.96	5.24
Time to event survival model					
$\alpha_0$	Health - MCI	0.010	0.002	0.007	0.014
	Health - Death	0.009	0.002	0.007	0.013
	MCI - Dementia	0.346	0.181	0.110	0.795
	MCI - Death	1.86	4.43	1.01	16.9
Family history	Health - MCI	-0.30	0.19	-0.68	0.08
	Health - Death	-0.32	0.19	-0.68	0.05
	MCI - Dementia	0.68	0.57	-0.45	1.78
	MCI - Death	-5.34	7.39	-21.5	7.89
APOE4	Health - MCI	0.069	0.190	-0.306	0.437
	Health - Death	0.049	0.190	-0.326	0.423
	MCI - Dementia	-0.084	0.555	-1.187	0.981
	MCI - Death	-8.27	6.55	-23.2	2.01
Sex	Health - MCI	-0.235	0.210	-0.652	0.168
	Health - Death	-0.229	0.187	-0.590	0.133
	MCI - Dementia	0.303	0.556	-0.770	1.411
	MCI - Death	-8.62	6.42	-23.17	1.43
Head injury	Health - MCI	-0.138	0.396	-0.956	0.601
	Health - Death	-0.006	0.241	-0.482	0.466
	MCI - Dementia	-1.12	1.48	-4.35	1.47
	MCI - Death	-6.92	7.53	-22.8	8.78
Sex*	Health - MCI	0.371	0.475	-0.555	1.306
Head injury	Health - Death	0.368	0.357	-0.325	1.039
	MCI - Dementia	1.322	1.610	-1.570	4.790
	MCI - Death	-0.327	9.896	-19.99	19.00

currently.

### 3.4.3 Piecewise constant hazard rates

Considering that time homogeneous assumption on hazard rate might be too restrictive, we divide the time into pieces by a set of pre-specified points,  $\tau_0, \dots, \tau_J$ , and assume constant intensity within each piece. Otherwise the model keeps the same as that in the last subsection. For simplicity, we assume the number of pieces  $J$  are the same for all transitions. So  $\alpha_{k0} = (\alpha_{k0}[1], \dots, \alpha_{k0}[J])$ , and

$$H_{k0,i} = \sum_{j < j_{k,i}} \alpha_{k0}[j] * (\tau_j - \tau_{j-1}) + \alpha_{k0}[j_{k,i}] * (t_{k,i} - \tau_{j_{k,i}-1})$$

where  $k = 1, 2, 3, 4$  and  $j_{k,i}$  tells which piece  $t_{k,i}$  falls in. The estimated values of intensities by piece are summarized in Table 3.7. Not surprisingly, both hazards of converting to MCI and death increase as age increases, and almost within every piece, the hazard rate of converting to MCI is higher than that of death. However, the hazard rate of death from MCI is almost zero, this could be due to the very limited follow-up years after converting to MCI. The other possibility is that subjects who converted to MCI convert to dementia before death, and dementia is a competing risk of death in this case.

### 3.4.4 Parametric hazard rates

Based on the piecewise hazard rates, the Weibull distributions were assumed for baseline survival functions. Thus, the cumulative hazard function has the form:

$$H_{k0,i} = \left( \frac{t_{k,i} - c_k}{b_k} \right)^{a_k},$$

where  $k = 1, 2, 3, 4$  and  $a_k, b_k$ , and  $c_k$  are Weibull parameters.  $c_k$  are set to be the smallest times observed for each transition, i.e.,  $c_1 = c_2 = 61, c_3 = c_4 = 0$ . Table 3.8

Table 3.7: Baseline hazard rates

Time	Health - MCI Mean hazard	Health - Death Mean hazard
0-60	0	0
60-65	0.00654	0.00009459
65-70	0.003649	0.002731
70-73	0.01288	0.009866
73-75	0.02548	0.01165
75-78	0.02312	0.01252
78-80	0.0307	0.03303
80-83	0.08775	0.04802
83-85	0.09602	0.05178
85-88	0.1781	0.09791
88-90	0.2919	0.1516
90-93	0.2361	0.2287
93-95	0.331	0.17
95-105	0.9607	0.3219
	MCI - Dementia	MCI - Death
0	0	0
0-2	0.1649	0
2-3	0.8257	0
3-9	3.549	0

shows the parameter estimates and their SDs and 95% confidence intervals for logistic models and survival models. Similar to the results from homogeneous intensities, APOE4 and family history are significant for predicting probability of converting to MCI from normal healthy, with APOE-4 allele present having higher hazard rate (hazard ratio =  $\exp(0.772)$ =2.16) and family history present having higher hazard rate (hazard ratio =  $\exp(0.495)$ =1.64). Besides these, the interaction between sex and head injury is also significant, notice that the head injury is almost significant, and these two factors have opposite effect. So the combination of two factors tells that head injury in males somehow is protective against converting to MCI. The hazard of converting to MCI for females with head injury history is about 3 times that of males with head injury history. None of the covariates included are significant for any survival models, although family history is close to being a significant factor for

time to dementia from MCI and sex is close to being a significant for time to MCI or death from health. Again, the large SDs in survival models for time to death or dementia from MCI is due to the very limited cases of transitions from MCI to death or dementia observed in the study currently.

In Figure 3.2, the black circles represent the hazard rates over time, resulting from piecewise transition intensity modeling, and the red curve represents the parametric hazard rate after assuming Weibull distribution survival function whose parameters are estimated from parametric transition intensity modeling. As we can see from the figure, the parametric hazard rate functions for MCI and death (from health) fit the piecewise hazards reasonably well.

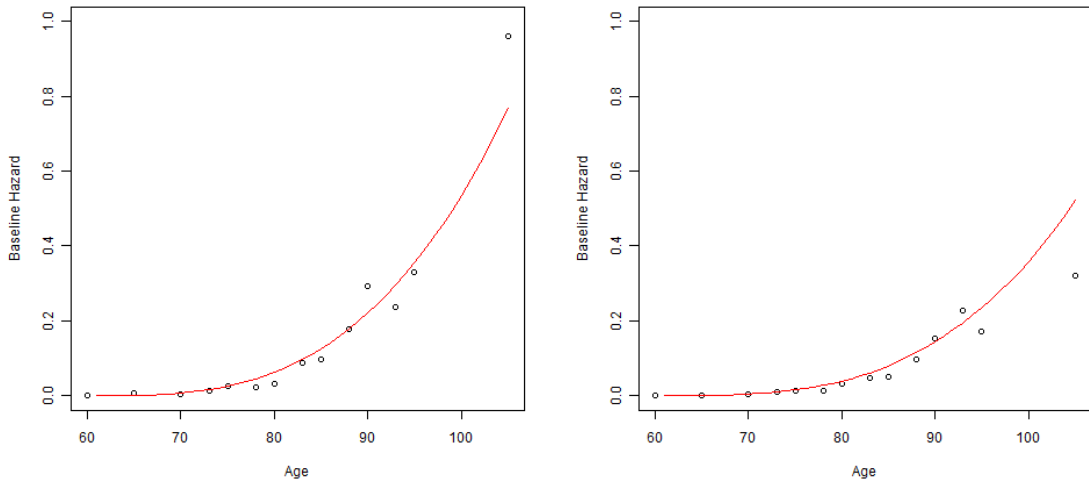


Figure 3.2: Baseline hazards

The random effect was further introduced into the parametric model with cumulative hazard function

$$H_{k,i} = H_{k0,i} \exp(\beta^T Z_i + b_i)$$

where  $H_{k0,i}$  is baseline cumulative hazard and  $b_i$  is the subject-specific random effect. As discussed in the next section, including such random effect doesn't fit the model any better.

Table 3.8: Parameter estimates under parametric hazard rates

Parameter	Transition type	Mean	SD	95% Lower	95% upper
Logistic modeling of probability of MCI					
Family history		0.495	0.245	0.015	0.985
APOE4		0.772	0.247	0.285	1.263
Sex		0.077	0.268	-0.453	0.615
Head injury		-0.819	0.439	-1.706	0.009
Sex*Head injury		1.099	0.563	0.008	2.208
Logistic modeling of probability of Dementia					
Family history		0.778	0.672	-0.493	2.183
APOE4		-1.016	0.662	-2.362	0.251
Sex		0.217	0.661	-1.12	1.517
Head injury		-0.703	1.800	-4.338	2.914
Sex*Head injury		1.279	2.047	-2.773	5.318
Time to event survival model					
a	Health - MCI	3.97	0.26	3.47	4.48
	Health - Death	4.11	0.23	3.65	4.57
	MCI - Dementia	1.99	0.45	1.20	2.96
	MCI - Death	6.69	2.26	2.00	9.86
b	Health - MCI	25.8	1.36	23.2	28.6
	Health - Death	29.4	1.15	27.2	31.7
	MCI - Dementia	3.27	0.95	2.00	5.69
	MCI - Death	31.7	11.8	8.50	49.2
Family history	Health - MCI	0.126	0.192	-0.245	0.490
	Health - Death	0.140	0.175	-0.207	0.471
	MCI - Dementia	1.086	0.590	-0.084	2.241
	MCI - Death	-3.342	8.583	-21.19	12.08
APOE4	Health - MCI	0.150	0.190	-0.227	0.524
	Health - Death	0.223	0.183	-0.134	0.586
	MCI - Dementia	0.084	0.575	-1.120	1.175
	MCI - Death	-4.324	8.047	-21.09	10.45
Sex	Health - MCI	-0.330	0.211	-0.763	0.091
	Health - Death	-0.237	0.180	-0.587	0.114
	MCI - Dementia	0.566	0.543	-0.455	1.642
	MCI - Death	-4.142	8.094	-20.99	10.44
Head injury	Health - MCI	0.010	0.393	-0.793	0.759
	Health - Death	0.132	0.240	-0.363	0.589
	MCI - Dementia	-1.161	1.594	-4.484	1.668
	MCI - Death	-2.545	8.642	-20.26	13.43
Sex* Head injury	Health - MCI	0.115	0.484	-0.832	1.076
	Health - Death	0.381	0.339	-0.298	1.030
	MCI - Dementia	1.163	1.740	-2.061	4.725
	MCI - Death	-0.307	9.716	-19.649	18.74



We then separate the transitions from health to MCI and the transitions from health to dementia. Instead of imputing transition types, we construct the likelihood through a multinomial distribution. More specifically, for each subject,  $(S_1, S_2, S_3, S_4) \sim \text{Multinomial}(P_{jj'}, 1)$ , where  $S_i$  is indicator for state  $i$  and  $P_{jj'}$  is element of the transition matrix. Table 3.9 summarizes the estimates and standard errors. APOE4 and head injury are significant for overall transition probability from health state to MCI. APOE4 and family history are significant for overall transition probability from health state to dementia. Only family history is significant for the duration time at MCI before converting to dementia. Weibull shape parameters for transitions from health(1) are significantly larger than 1, suggesting time dependent hazard rates. However, the Weibull shape parameters for transitions from MCI(2) are not significantly different from 1, indicating time homogeneous hazard rates. Therefore, we apply a semi-Markov and Markov mixture model to the BRAiNS data, assuming a semi-Markov process from transitions from health state and a Markov process for transitions from MCI. The corresponding estimates are included in Table 3.10. The parameter estimates are similar to those obtained from pure semi-Markov model except that family history is not significant for transition intensity from MCI to dementia.

### 3.4.5 Model comparison

The Markov model estimated using “msm” package and the other models estimated by assuming a mixture modeling of transition probabilities and hazard rates provide different interpretations. In intensity only Markov model, family history, APOE4, and sex by head injury interaction are significant factors affecting some transition intensities. However, none of these factors are significant for any hazards in the mixture models. But these factors are significant in predicting transition probabilities in the mixture models.

Table 3.9: Parameter estimate(standard error) from pure semi-Markov model

Component	Tran	a	b / int	apoe4	famhx	sex	headinj	sex*headinj
Overall	1 - 2		-0.89(0.27)	0.70(0.28)	0.36(0.27)	-0.29(0.29)	-0.98(0.50)	1.3(0.6)
Transition	1 - 3		-2.40(0.40)	0.81(0.35)	0.75(0.35)	0.82(0.44)	-0.56(0.76)	0.82(0.89)
Probability	2 - 3		0.71(0.75)	-1.00(0.73)	0.30(0.78)	-0.03(0.80)	-0.06(1.97)	0.32(2.23)
Survival	1 - 2	4.22(0.35)	26.1(1.4)	0.22(0.24)	0.11(0.25)	-0.08(0.25)	-0.05(0.48)	-0.50(0.61)
Time	1 - 3	4.70(0.51)	32.7(3.7)	0.49(0.38)	0.14(0.33)	-0.04(0.49)	0.58(0.86)	0.42(0.93)
	1 - 4	4.17(0.23)	30.4(1.2)	0.21(0.18)	0.16(0.17)	-0.23(0.18)	0.12(0.23)	0.44(0.34)
	2 - 3	1.71(0.43)	4.70(2.1)	-0.11(0.62)	1.4(0.6)	0.71(0.60)	-1.3(1.5)	1.7(1.6)
	2 - 4	0.87(0.27)	8.6(8.8)	-0.45(0.80)	-0.73(0.82)	-0.22(0.90)	-0.48(1.65)	1.6(2.0)

Table 3.10: Parameter estimate(standard error) from semi-Markov + Markov mixture model

Component	Tran	a	b / int	apoe4	famhx	sex	headinj	sex*headinj
Transition	1 - 2		-0.89(0.27)	0.69(0.28)	0.36(0.28)	-0.29(0.30)	-0.97(0.50)	1.23(0.66)
Probability	1 - 3		-2.40(0.43)	0.80(0.35)	0.75(0.35)	0.81(0.44)	-0.58(0.77)	0.84(0.89)
Survival	1 - 2	4.23(0.35)	26.1(1.5)	0.21(0.25)	0.12(0.25)	-0.08(0.26)	-0.06(0.48)	-0.49(0.61)
Time	1 - 3	4.69(0.52)	32.8(3.7)	0.49(0.38)	0.15(0.32)	-0.03(0.47)	0.56(0.86)	0.44(0.94)
	1 - 4	4.16(0.23)	30.3(1.1)	0.20(0.18)	0.16(0.17)	-0.24(0.18)	0.11(0.23)	0.45(0.34)
Transition	2 - 3		0.165(0.058)	-0.72(0.46)	0.65(0.43)	0.25(0.43)	-1.37(1.26)	1.86(1.41)
Intensity	2 - 4		0.091(0.043)	0.38(0.62)	-0.46(0.69)	-0.27(0.65)	-0.77(1.30)	1.61(1.56)

These two types of models used different likelihood functions as discussed in Section 1, so there is no test available to compare them. As stated earlier, we think in the BRAiNS data, the latter is likely a better choice. We therefore focus now on comparing models within the Bayesian method.

The estimated DIC values for the fitted models, where conversions to MCI and dementia from health, are combined are summarized in Table 3.11. Clearly, the model with time-homogeneous transition intensities has the largest DIC, which indicates the transition intensities are time-dependent. The time here is actually duration in previous state. The model with parametric intensities fit the data well. However, the parametric model that includes random effect doesn't fit the model any better.

Table 3.11: Model comparison using DIC

Model	DIC
Time homogeneous transition intensities	4976.4
Piecewise constant transition intensities	NA
Parametric transition intensities	3423.7
Parametric transition intensities with shared random effect	3444.4

A couple of independent runs were carried for each Bayesian model. Every 50th samples were collected after discarding first 10000 sample in each chain with total number of samples collected about 3600. The upper limits for 'potential scale reduction factor', calculated by Gelman and Rubin method, are close to 1 for all parameters, which means the approximate convergence is achieved. The sampling history and auto-correlation for each parameter are used to check the sampling efficiency. Since all parameters show similar pattern, we only show the plots for coefficients of family history here (Figure 3.3 and Figure 3.4). The sampling history (time series) plots look like white noise, and the chain is well mixing. The auto-correlation plot show the correlations are close to 0 after lag number  $\geq 2$  and thus the samples are approximately independent.

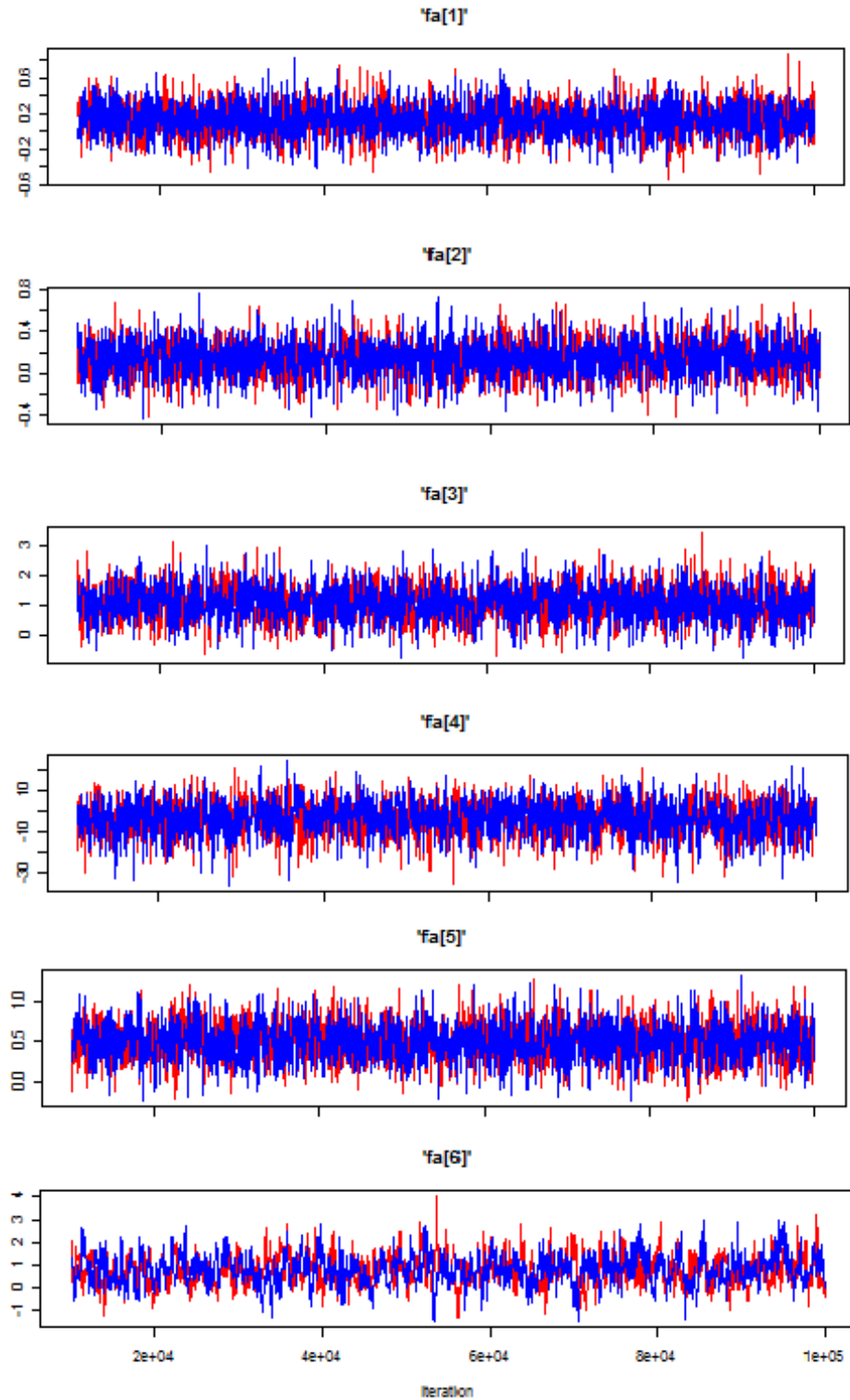


Figure 3.3: Sampling history of coefficient for family history. Indices 1-4 represent 4 transition types and indices 5-6 represent probability models

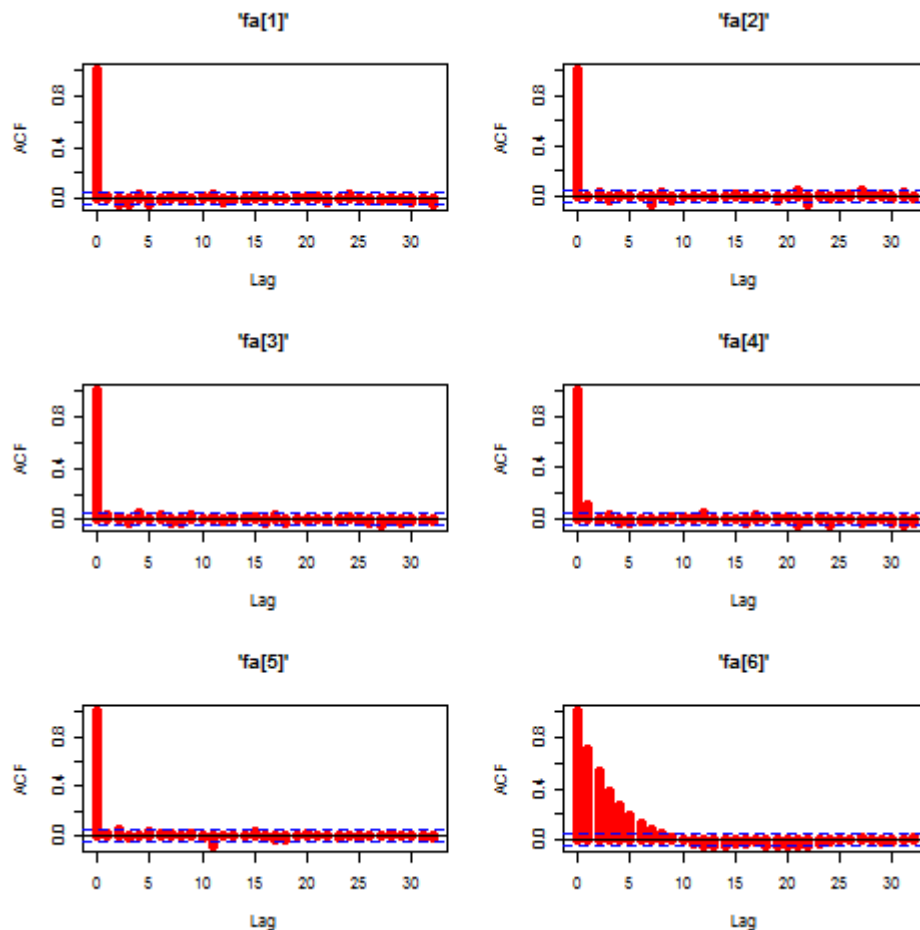


Figure 3.4: Autocorrelation of coefficient for family history. Indices 1-4 represent 4 transition types and indices 5-6 represent probability models

### 3.5 Discussion

In this Chapter, we introduced two methods of likelihood function construction for panel data. One based on transition intensity matrix and assumed a Markovian process, and the other based on mixture modeling of transition probabilities and hazard rates with semi-Markov assumption. In BRAiNS study, we think that some healthy subjects with particular covariates are more likely to develop MCI before death, while some may be very unlikely to develop MCI. Thus, it is more appropriate to distinguish these two groups before investigating their hazards of converting to MCI. We

have proposed a competing risks type Bayesian mixture model with some complexity, i.e. many unknown parameters especially piecewise hazard rates were assumed, and data imputation, but simple implementation with the help of OpenBUGS, especially R package 'BRugs' in R. This model has several advantages.

First, compared to the traditional Markov models which are totally transition intensities specified, our model takes the mixture structure of probability and hazards, this takes into account the following fact: When we are investigating the hazard of converting to MCI from normal health, basically we are looking at the ages when subjects become MCI, and treating death from health as right censored observations. If subjects who died before converting to MCI are those who will never experience MCI, then the death times tend to over-estimate the times to MCI. Distinguishing these two events will eliminate such bias.

Second, Bayesian modeling allows the inclusion of additional terms, for example, random effects. Although including random effects in our proposed model doesn't fit the data any better. This may be due to that we specify the same random effect for all transition intensity regressions. The drawback with including additional parameters is more computation time is needed in MCMC simulation. The model with non-parametric (or piecewise) baseline transition intensities requires specification of most parameters and takes much longer time. However, we found that Weibull distribution is a good fit to the baseline survival function and this reduces a large amount of unknown parameters to save time in computation.

Third, the mixture model facilitate the data augmentation. In a Bayesian approach, data augmentation is performed through data imputation in each MCMC simulation step and thus makes data complete. In our modeling of BRAiNS study, for subjects who are alive and healthy, the types of event from normal health are unknown, but in each step of MCMC simulation, types will be imputed based on the logistic model we specified for probability of converting to MCI, and further be used in the likelihood

computation in the survival part of the model.

Fourth, with OpenBUGS's zero trick, any model will work as long as there is a likelihood specified. This allows us to incorporate the interval censored data. We could specify the contributed likelihood for exact observed times, right censored or interval censored observations.

The model seems to work better only when there is an adequate number of transitions happening. Since the large number of events fits the probability model better which in turn will impute the unknown event types more close to the real data. In the current BRAiNS dataset, there are quite limited cases of converting from MCI to dementia or to death. However, this won't be an issue once more years of follow-up are realized for those subjects who converted to MCI from health.

In summary, we found that APOE4, family history, and sex by head injury interaction are significant factors for predicting probability of converting to MCI from normal healthy, with APOE-4 allele present, family history present, and females who had head injury having higher hazard rate. But none of the covariates interested are significant for any survival models.

Although it is reasonable to utilize a mixture model for the transitions from normal health state, with multinomial logistic model for the overall transition probabilities and survival models for time to events given specific transitions. It might not be good idea to apply the mixture model for the transitions from MCI state. Since clinical MCI behaves as biomarker for dementia, so eventually all clinical MCI will convert to dementia and the probability model does not work in this case. Instead, treating death from MCI as right censored observations when times to dementia (from MCI) is our interest. Alternatively, we assumed a Markov process for the transitions from MCI. The resulting model is a mixture of semi-Markov and Markov processes. The likelihood was constructed through a multinomial distribution.



## Chapter 4 Discussions and Future Research

In summary, we proposed a direct modeling of CIF for competing risks data without assuming dependence among risks. The model accommodates the left censored data and ensure that the sum of the CIFs never exceeds one. We also proposed a semi-Markov multi-state model by modeling the overall transition probability and duration time distribution. The model allows interval censored data and could incorporate the random effects.

Since under non-informative censoring, the likelihood function for competing risks data factors as a product of individual likelihoods, one for each event. Thus the basic simple models of competing risks data based on the CSHs could be analyzed by a series of standard survival analyses, one risk at a time, treating failures from other risks as right censored observations. These simple models have no common parameters for any two or more CSH functions, i.e., there are no covariates for which the effect is assumed to be the same on several CSHs. Notice also as in Cox regression models, some parameters may be forced to be common for several risks.

However, there is a key difference between such analyses on CSHs and a standard survival analysis. The simple relation of survival probability and hazard function in standard survival analysis no longer holds in the competing risks framework. The transformation of the CSH through  $1 - \exp(-\text{cumulative CSH})$  does not have a probability interpretation. Although under an assumption of independent competing risks, the CIF, which is well defined and has probability interpretation, can be obtained by the above transformation. So basically in competing risks analyses, CSHs model are assumed and estimated cumulative incidence functions are computed based on these CSHs. A drawback to this approach, as pointed out by Gray (1988) and Pepe (1991), is that CIFs do not depend on the covariates in the same way as CSHs do. Because

of these concerns, Fine and Gray (1999) introduced a new semi-parametric competing risks regression model where the CIFs are directly regressed on the covariates of interest. Jeong and Fine (2006) proposed parametric modeling of CIFs and found that using an improper Gompertz distribution as baseline distribution is a good fit to their data. Such parametric model is able to handle censoring data.

Both CSH regression models and CIF regression models are created based on the same likelihood function. The difference is that the likelihood expressed in terms of CIFs does not factor into separate pieces, one for each risk. Under this formulation, specification of one CIF might affect the other CIFs. It acts like introducing dependence among risks. We have applied such parametric modeling of CIFs to the Nun study. The link function,  $\log\{-\log(1 - F)\}$ , where  $F$  represents a CIF, was used since it corresponds to a proportional SH model (Jeong and Fine, 2007). The resulting parameters are comparable to those obtained by CSH modeling. The reason is that their likelihood functions are the same.

These direct models of CIFs used improper distributions, either parametric form or non-parametric form, to model the CIFs. However, Jeong and Fine did not consider the constraint that all marginal risk probabilities add to 1 under any covariate configurations in their models and the resulting CIFs could sum to a number, greater than 1, which is an error. In our proposed parametric model, we took such a constraint into account, assuming a proportional SH for the risk of interest and another similar but no longer proportional SH model for other risks. The marginal probability of risk of interest is also a parameter in the model.

As mentioned above, the mis-specification of one risk might effect others, the correct improper baseline parametric distribution should be used to capture the features of the data. Weibull distribution, which graphically fits the non-parametric MLEs of CIFs quite well in the Nun study, was used to parameterize the baseline distribution. Note also that Weibull distribution could well fit a CIF with or without a plateau,

which enables it to be applied for other types of competing risks data. It turned out that the parameters for the risks of interest are similar to those that would be obtained from Jeong's proportional SH model, but the parameters corresponding to other risks are not comparable to those obtained from Jeong's model. The disadvantage of our proposed model is that the interpretation of the parameters for the other risks is not straightforward. The fit of Weibull distribution the CIF is examined by comparing the two graphically. It is also interesting to know whether the CSH model or direct model of CIF fits the data better. One possible future work in this area is to propose a more specific goodness-of-fit test.

Considering that dementia and death are actually forming a semi-competing risks data, i.e., the subject who converted to dementia will further experience death, but the subjects who died will never encounter dementia. In such situation, the death times are relatively larger than times to dementia, hence treating death times as censored when dementia is the risk of interest will overestimate the survival times to dementia (or underestimate the hazard rate of dementia). This is more obvious, when clinical MCI, a phase before dementia, together with death are the two competing risks, as discussed in Chapter 3. Another fact is that some subjects who are more likely to convert to MCI converted at relatively older ages. With these considerations, a mixture model, as first introduced by Larson and Dinse (1985) is more promising in semi-competing risks data. Mixture models which jointly model event probabilities and survival times were utilized in multi-state modeling in Chapter 3.

Transition intensities and probabilities are the most important quantities in multi-state models. Simple models assume a Markov model by characterizing the transition probabilities, for example, through a multinomial logistic regression (Salazar, et al., 2007). More general models are based on transition intensities. The Bayesian version of modeling transition probabilities is assuming a multinomial distribution. Each transition probability is expressed in terms of transition intensity using the Kol-

mogorov equation (Pan, et al., 2007), and covariate effects can be specified through the regression models for transition intensities. The advantage of the Bayesian approach in this situation is to take the time intervals into account. Such Markov modeling of probability with the consideration of visit times, as well as other covariates might be worth further investigation.

The likelihood function is always constructed based on transition probabilities. Modeling transition intensities become difficult when the process is non-homogeneous Markov or non-Markov because transition probabilities are no longer easy to obtain from the transition intensities. Therefore, modeling transition probabilities directly is good alternative since it avoids calculating transition probability from the intensities. We proposed a semi-Markov multi-state model by directly modeling transition probability through a multinomial distribution. The transition probability consists two components: one for overall transition probability, the other for conditional survival probability. The computations were done with the help of OpenBUGS software and R package “BRugs”. The Bayesian model allows complex structure, i.e., hierarchical models. The model will work if random effects are included. It is easy to obtain both point estimate and uncertainty of specified function using Bayesian approach. The other advantages of Bayesian modeling through MCMC simulation technique is handling missing values through data augmentation.

It turns out that cumulative hazard functions resulting from Weibull distribution with three parameters fit the cumulative hazard functions obtained from piecewise hazard quite well. The Weibull shape parameter is close to 1 for transitions from MCI to dementia or to death, suggesting a Markov process. Thus, We consider both constant transition intensities and duration time dependent transition intensities, leading to a mixture of Markov and semi-Markov processes, and model both transition intensities and transition probabilities. Including the shared random effect in the model did not improve the fitting, based on DIC. As a future work, we may specify different

magnitudes of random effects for different types of transitions, but these effects are correlated. As a Bayesian model, it is always of interest to know the effect of priors on the estimation. The missing transition times can be imputed through data augmentation in Bayesian approach.

As discussed in Chapter 3, the four-state model of BRAiNS data consists of two sets of (semi-)competing risks data. However, these two are different because of the following: clinical MCI behaves as biomarker for dementia, so eventually all clinical MCI will convert to dementia if they would live long enough. By this we mean for those who died before converting to dementia from MCI are going to convert to dementia if they would live longer. If this is the case, treating death from MCI as right censored observations when times to dementia (from MCI) is our interest might be more appropriate.

Multi-state model is the most general way describing longitudinal failure time data. Depending on the research question, sometimes maybe a portion of the whole data is considered, for example, competing risks data. Another question of interest is “does converting to MCI shorten the life expectancy?”. In this case, there is only one risk, which is death. The data becomes standard survival data and the effect of converting to MCI acts as an effect of a time-dependent covariate. Similar effects can be introduced into our proposed multi-state models through modeling overall transition probabilities and through modeling conditional survival probability.

## Appendices

### I. Basic quantities in competing risks modeling

#### a. Overall hazard rate $\lambda(t)$

$$\lambda(t) = \lim_{\Delta \rightarrow 0} \frac{P[t \leq T \leq t + \Delta | T \geq t]}{\Delta}$$

#### b. Cause-specific hazard rate $\lambda_j(t)$

$$\lambda_j(t) = \lim_{\Delta \rightarrow 0} \frac{P[t \leq T \leq t + \Delta, J = j | T \geq t]}{\Delta}$$

Thus,

$$\lambda(t) = \sum_{j=1}^J \lambda_j(t)$$

#### c. Overall cumulative hazard $\Lambda(t)$

$$\Lambda(t) = \begin{cases} \int_0^t \lambda(s) ds, & \text{continuous} \\ \sum_{s \leq t} \lambda(s), & \text{discrete} \end{cases}$$

#### d. Empirical survivor function $F(t)$

$$F(t) = P[T > t] = \begin{cases} \exp(-\Lambda(t)), & \text{continuous} \\ \prod_{t_i \leq t} (1 - d\Lambda(t_i)), & \text{discrete} \end{cases}$$

#### e. Cumulative incidence function $F_j(t)$

$$F_j(t) = P[T \leq t, J = j]$$

For discrete models, the jump values of the sub-distribution is

$$p_j(t) = P[T = t, J = j] = F_j(t) - F_j(t-)$$

## II. A preliminary draft of R code for computing NPMLEs

To use this code, input the variables T (time), DELTA (1 for observed events; 0 for right censored, -1 for left censored) D1(indicator for experiencing event 1), and D2 (indicator for experiencing event 2). CIF <- function (data0, iternum, eps) {

```
t.min <- min( data0$T[data0$DELTA==1] )
```

```
data0 <- data0[which(data0$T>=t.min),]
```

```
#sort the data by t
```

```
data1 <- data0[order(data0$T, na.last=NA) , ] #NA are removed.
```

```
t <- data1$T
```

```
delta <- data1$DELTA
```

```
d1 <- data1$D1
```

```
d2 <- data1$D2
```

```
type <- d1+2*d2
```

```
#calculate the followings: e1, e2, te, tl, tc, c, nt, nc, ne, nr
```

```
Nobs <- length(t) #Nobs:number of total subjects
```

```
n0 <- length(t[delta==1]) #n0:number of events
```

```
te <- unique(t[delta==1]) #distinct event times
```

```
ne <- length(te) #ne:number of distinct event times
```

```
tl <- t[delta==-1] #vector recording left-censored times for type 1 event
```

```
nl <- length(tl) #nl:number of left times
```

```
lt <- type[delta==-1] #lt: left-censored type indicator
```

```
tc <- t[delta==0]
```

```
nc <- length(tc)
```

```
nr <- vector(length=ne) #vector of number at risk
```

```
c <- vector(length=ne) #vector of right censored obs
```

```
e1.0 <- tapply(d1[delta==1],t[delta==1],sum)
```

```
e2.0 <- tapply(d2[delta==1],t[delta==1],sum)
```

```

nt.0 <- e1.0 + e2.0 #number of events at each time point (complete data)
for (i in 1:ne) { c[i] <- 0
for (j in 1:nc) { if( (i<ne & (te[i]<=tc[j] & tc[j]<te[i+1]))
— (i==ne & te[i]<=tc[j]) ) c[i]<- c[i]+1 }}
nr[1] <- Nobs - nl
for (i in 2:ne) { nr[i] <- nr[i-1] - nt.0[i-1] - c[i-1] }
#cause-specific hazard rate
lam1 <- e1.0/nr
lam2 <- e2.0/nr
lam <- lam1 + lam2
p <- vector(length=ne)
p1 <- vector(length=ne)
p2 <- vector(length=ne) #point mass (p1,p2,p)
F <- vector(length=ne)
F1 <- vector(length=ne)
F2 <- vector(length=ne)
LAM <- vector(length=ne)
LAM1 <- vector(length=ne)
LAM2 <- vector(length=ne)
F_new <- vector(length=ne)
F1_new <- vector(length=ne)
F2_new <- vector(length=ne)
F[1] <- 1-lam[1]
p[1] <- lam[1]
p1[1] <- lam1[1]
p2[1] <- lam2[1]
for (k in 1:ne) {

```



```

LAM1[k] <- sum(lam1[1:k])
LAM2[k] <- sum(lam2[1:k])
LAM[k] <- LAM1[k] + LAM2[k]
if (k>1) {F[k] <- F[k-1] * (1-lam[k])
p[k] <- F[k-1] - F[k]
p1[k] <- lam1[k] * F[k-1]
p2[k] <- lam2[k] * F[k-1]} }
for (k in 1:ne) { F1[k] <- sum(p1[1:k])
F2[k] <- sum(p2[1:k])}
plot(te,F2,lty=1, type="s")
lines(te,F1,lty=1, type="s")
for (iter in 1:iternum) {
#update e1, e2 and nt
d1_delta <- matrix(rep(0,ne*nl), nrow=ne, ncol=nl)
d2_delta <- matrix(rep(0,ne*nl), nrow=ne, ncol=nl)
for (j in 1:nl) {
if ( max(F1[te<=tl[j]]) >0 )
d1_delta[,j] <- (p1/max(F1[te<=tl[j]])) * ( as.numeric(te<=tl[j]))
* ( as.numeric(lt[j]==1))
if ( max(F2[te<=tl[j]]) >0 )
d2_delta[,j] <- (p2/max(F2[te<=tl[j]])) * ( as.numeric(te<=tl[j]))
* ( as.numeric(lt[j]==2))
}
e1 <- e1_0 + apply(d1_delta,1,sum)
e2 <- e2_0 + apply(d2_delta,1,sum)
nt <- e1+e2 #number of events including left censoring contributions
#update nr

```

```

nr[1] <- Nobs
for (i in 2:ne) { nr[i] <- nr[i-1] - nt[i-1] - c[i-1] }
#update cause-specific hazard rate
lam1 <- e1/nr
lam2 <- e2/nr
lam <- lam1 + lam2
#update F, F1 and F2
F_new[1] <- 1-lam[1]
p[1] <- lam[1]
p1[1] <- lam1[1]
p2[1] <- lam2[1]
for (k in 1:ne) {
LAM1[k] <- sum(lam1[1:k])
LAM2[k] <- sum(lam2[1:k])
LAM[k] <- LAM1[k] + LAM2[k]
if (k>1) {F_new[k] <- F_new[k-1] * (1-lam[k])
p[k] <- F_new[k-1] - F_new[k]
p1[k] <- lam1[k] * F_new[k-1]
p2[k] <- lam2[k] * F_new[k-1]} }
for (k in 1:ne) { F1_new[k] <- sum(p1[1:k])
F2_new[k] <- sum(p2[1:k])}
if (sum(abs(F1_new-F1))<eps & sum(abs(F2_new-F2))<eps)
break
F =F_new
F1=F1_new
F2=F2_new }
list(te, F1, F2, F, iternum=iter)

```

```

plot(te,F2,lty=1, type="s")
lines(te,F1,lty=1, type="s")
}
test <- CIF(nun, iternum=100, eps=1e-10)

```

### III. SAS Code for parametric competing risks modeling

```

PROC nlp DATA=nun tech=newwrap outest=parms COV=2 PCOV;
eps=0.001;
MAX LogLik;
PARMS lam_1=16.3, alpha_1=2.65, c_dem=0.5, c_1=75.52, c_2=74.90,
lam_2=16.97, alpha_2=3.08, beta1_apoe4=0.2, beta2_apoe4=0.1,
beta1_high=-0.05, beta2_high=0.2, beta1_low=0.05, beta2_low=0.1;
BOUNDS c_dem<1;

exp_beta1=exp(beta1_apoe4*apoe4+beta1_high*high+beta1_low*low);
exp_beta2=exp(beta2_apoe4*apoe4+beta2_high*high+beta2_low*low);

shift1=(t>c_1)*(t-c_1)+(t<c_1)*eps;
shift2=(t>c_2)*(t-c_2)+(t<c_2)*eps;

F1=1-exp( -(shift1/lam_1)**alpha_1 );
F2=1-exp( -(shift2/lam_2)**alpha_2 );

CDF1=1-( 1 - c_dem + c_dem*exp( -(shift1/lam_1)**alpha_1 ) )**exp_beta1;
pdf1=(1-c_dem*F1)**exp_beta1*(exp_beta1)*(1/(1-c_dem*F1))*c_dem
*(alpha_1/lam_1)*((shift1/lam_1)**(alpha_1-1))*exp(-(shift1/lam_1)**alpha_1);

```

```

CDF2=(1-c_dem)**exp_beta1*(1-exp(-exp_beta2*(shift2/lam_2)**alpha_2));
pdf2=(1-c_dem)**exp_beta1*exp(-exp_beta2*(shift2/lam_2)**alpha_2)*exp_beta2
*alpha_2/lam_2*(shift2/lam_2)**(alpha_2-1);

if 1-CDF1-CDF2>0 then log_surv=log(1-CDF1-CDF2); else log_surv=log(eps) - 1.5
+ 2*(1-CDF1-CDF2)/eps - 0.5*(((1-CDF1-CDF2)/eps)**2);

LogLik = (left=0)*( (delta=1)*log(pdf1)+(delta=2)*log(pdf2)
+(delta=0)*log_surv ) + (left=1)*log(CDF1);

RUN;

```

#### IV. R code for Markov model using “msm” package

```

library(msm)
library(foreign)
brains <- read.xport("brains_msm.xpt")
sink("stat.txt", append=FALSE, split=FALSE) #output directed to stat.txt
statetable.msm(STATE, PTID, data = brains) #transition numbers
#assign initial transition intensities
twoway4.q <- rbind(c(0,0.034,0,0.04),c(0,0,0.3,0.09),c(0,0,0,0),c(0,0,0,0))
#1. Panel data without covariates
msm0 <- msm(STATE~T,subject=PTID,data=brains,qmatrix=twoway4.q,death=4)
pmatrx.msm(msm0, t = 5, ci = "normal")
msm <- msm(STATE ~ T, subject = PTID, data = brains, qmatrix = twoway4.q,
pci=c(5, 8, 10, 15), death = 4)

```

```

lrtest.msm(msm0, msm)

#2. Panel data with covariate, including baseline age
cov.msm <- msm(STATE ~ T, subject = PTID, data = brains, covariates = ~ T0 +
FAMHX + APOE4 + SEX_I+HEADINJ+SEXHEAD, qmatrix = twoway4.q, death
= 4, method = "BFGS", control = list(fnscale = 4000, maxit = 10000))
hazard.msm(cov.msm) #hazard ratios
lrtest.msm(msm0, cov.msm)

```

## V. R code for running Bayesian inference through MCMC using OpenBUGS

```

model
{
C <- 100
# Model
for(i in 1:N) {
zeros[i] <- 0
# Survival densities
xbeta1 <- ap[1]*apoe[i]+fa[1]*famhx[i]+se[1]*sex[i]
+hd[1]*headinj[i]+sh[1]*sexhead[i]
H10_l[i] <- pow((t1L[i]-c[1]+eps)/b[1],a[1])
H10_u[i] <- pow((t1U[i]-c[1]+eps)/b[1],a[1])
H1_l[i] <- H10_l[i] * exp(xbeta1)
H1_u[i] <- H10_u[i] * exp(xbeta1)
F1_l[i] <- 1 - exp(-H1_l[i])
F1_u[i] <- 1 - exp(-H1_u[i])
f1_u[i] <- exp(-H1_u[i])*exp(xbeta1) *a[1]/b[1]*pow((t1U[i]-c[1]+eps)/b[1],(a[1]-1))

```

```

xbeta2 <- ap[2]*apoe[i]+fa[2]*famhx[i]+se[2]*sex[i]
+hd[2]*headinj[i]+sh[2]*sexhead[i]
H20_l[i] <- pow((t2L[i]-c[2]+eps)/b[2],a[2])
H20_u[i] <- pow((t2U[i]-c[2]+eps)/b[2],a[2])
H2_l[i] <- H20_l[i] * exp(xbeta2)
H2_u[i] <- H20_u[i] * exp(xbeta2)
F2_l[i] <- 1 - exp(-H2_l[i])
F2_u[i] <- 1 - exp(-H2_u[i])
f2_u[i] <- exp(-H2_u[i])*exp(xbeta2) *a[2]/b[2]*pow((t2U[i]-c[2]+eps)/b[2],(a[2]-1))

```

```

xbeta3 <- ap[3]*apoe[i]+fa[3]*famhx[i]+se[3]*sex[i]
+hd[3]*headinj[i]+sh[3]*sexhead[i]
H30_l[i] <- pow((t3L[i]-c[3]+eps)/b[3],a[3])
H30_u[i] <- pow((t3U[i]-c[3]+eps)/b[3],a[3])
H3_l[i] <- H30_l[i] * exp(xbeta3)
H3_u[i] <- H30_u[i] * exp(xbeta3)
F3_l[i] <- 1 - exp(-H3_l[i])
F3_u[i] <- 1 - exp(-H3_u[i])
f3_u[i] <- exp(-H3_u[i])*exp(xbeta3) *a[3]/b[3]*pow((t3U[i]-c[3]+eps)/b[3],(a[3]-1))

```

```

xbeta4 <- ap[4]*apoe[i]+fa[4]*famhx[i]+se[4]*sex[i]
+hd[4]*headinj[i]+sh[4]*sexhead[i]
H40_l[i] <- pow((t4L[i]-c[4]+eps)/b[4],a[4])
H40_u[i] <- pow((t4U[i]-c[4]+eps)/b[4],a[4])
H4_l[i] <- H40_l[i] * exp(xbeta4)
H4_u[i] <- H40_u[i] * exp(xbeta4)

```

```

F4_l[i] <- 1 - exp(-H4_l[i])
F4_u[i] <- 1 - exp(-H4_u[i])
f4_u[i] <- exp(-H4_u[i])*exp(xbeta4) *a[4]/b[4]*pow((t4U[i]-c[4]+eps)/b[4],(a[4]-1))

# event probabilities D1[i]=1 or 2
# D1[i]=1 for MCI to Dementia, D2[i]=2 for MCI to Death
D1[i] ~ dcat(P1[i,])
P1[i,1] <- p1[i]
P1[i,2] <- 1-p1[i]
logit(p1[i]) <- int[1] + ap[5]*apoe[i] + fa[5]*famhx[i] + se[5]*sex[i]

D2[i] ~ dcat(P2[i,])
P2[i,1] <- p2[i]
P2[i,2] <- 1-p2[i]
logit(p2[i]) <- int[2] + ap[6]*apoe[i] + fa[6]*famhx[i] + se[6]*sex[i]
logl1[i,1] <- ((1-d1[i])*(-H1_u[i])
+d1[i]*log((1-I1[i])*f1_u[i]+I1[i]*(F1_u[i]-F1_l[i])+eps))
logl1[i,2] <- ((1-d2[i])*(-H2_u[i])
+d2[i]*log((1-I2[i])*f2_u[i]+I2[i]*(F2_u[i]-F2_l[i])+eps))
logl2[i,1] <- ((1-d3[i])*(-H3_u[i])
+d3[i]*log((1-I3[i])*f3_u[i]+I3[i]*(F3_u[i]-F3_l[i])+eps))*Tr3[i]
logl2[i,2] <- ((1-d4[i])*(-H4_u[i])
+d4[i]*log((1-I4[i])*f4_u[i]+I4[i]*(F4_u[i]-F4_l[i])+eps))*Tr4[i]

# Log Likelihood
logl[i] <- logl1[i,D1[i]]+logl2[i,D2[i]]

```

```

    # zeros trick of sampling
    phi[i] <- C-logl[i]
    zeros[i] ~ dpois(phi[i])
  }

  c[1] <- c1
  c[2] <- c2
  c[3] <- c3
  c[4] <- c4
  int[1] ~ dnorm(0.0,0.001)
  int[2] ~ dnorm(0.0,0.001)

  for (j in 1:4) {
    a[j] ~ dunif(1,10)
    b[j] ~ dunif(1,50)
    hd[j] ~ dnorm(0.0,0.01)
    sh[j] ~ dnorm(0.0,0.01)
  }
  for (j in 1:6) {
    ap[j] ~ dnorm(0.0,0.01)
    fa[j] ~ dnorm(0.0,0.01)
    se[j] ~ dnorm(0.0,0.01)
  }
}

```



## Bibliography

- [1] Aalen, O. (1976). Nonparametric Inference in Connection with Multiple Decrement Models. *Scandinavian Journal of Statistics* **3**, 15-27.
- [2] Aalen, O., Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* **5**, 141-150.
- [3] Abner, E.L., Cooper, G.E., Fardo, D., Jicha, G.A., Mendiondo, M.S., Nelson, P.T., Smith, C.D., Van, E. L.J., Wan, L., Schmitt, F.A., Kryscio, R.J., Role of risk factors and clinical mild cognitive impairment in transitions to dementia, in preparation.
- [4] Albert, A. (1962). Estimating the infinitesimal generator of a continuous time finite state Markov process. *Annals of Mathematical Statistics* **33**, 727C753.
- [5] Andersen, P. K., Klein, J., Rosthøj, S. (2003). Generalised linear models for correlated pseudo-observations with applications to multi-state models. *Biometrika* **90**, 15-27.
- [6] Andersen, P. K., Perme, M. P. (2008). Inference for outcome probabilities in multi-state models. *Lifetime Data Anal* **14**, 405-431.
- [7] Bladt, M., Sorensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *Journal of the Royal Statistical Society, Series B* **67**, 395C410.
- [8] Buckley J.D. (1984). Additive and multiplicative models for relative survival rates. *Biometrics*. **40**, 51-62.
- [9] Chen, M.-H., Shao, Q.-M., and Ibrahim, J.G. (2000). Monte Carlo methods in Bayesian computation, New York: *Springer*.
- [10] Commenges, D., Joly, P., Letenneur, L., Dartigues, J. F. (2004). Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in Medicine* **23**, 199-210.
- [11] Commenges, D., Joly, P., Gegout-Petit, A., Liqueur, B. (2007). Choice between semi-parametric estimators of Markov and non-Markov multi-state models from coarsened observations. *Scandinavian Journal of Statistics* **34**, 33-52.
- [12] Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological* **34**, 187-220.
- [13] Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association*, **94**, 496-509.

- [14] Fine, J. P. (2001). Regression modeling of competing crude failure probabilities. *Biostatistics* **2**, 85-97.
- [15] Foucher, Y., Giral, M., Soullou, J. P., Daures, J.P. (2010). A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. *Statistical Methods in Medical Research* **19**, 127-145.
- [16] Gentleman, R.C., Lawless, J.F., Lindsey, J.C., Yan, P. (1994). Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine* **13**, 805-21.
- [17] Gray, R. J. (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics* **16**, 1141-54.
- [18] Heckman, J. J. and Honore, B. E. (1989). The Identifiability of the Competing Risks Model, *Biometrika* **76**, 325-30.
- [19] Henderson, R., Diggle, P., Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.
- [20] Hennerfeind, A., Brezger, A., Fahrmeir, L. (2006). Joint modelling of longitudinal measurements and event time data. *Journal of the American Statistical Association* **101**, 1065-1075.
- [21] Hubbard, R.A., Inoue, L.Y., Fann, J.R. (2008). Modeling nonhomogeneous Markov processes via time transformation. *Biometrics* **64**, 843-50.
- [22] Jackson, C. H. (2011). Multi-State Models for Panel Data: The msm Package for R, *Journal of Statistical software* **38**.
- [23] Jeong, J. H. and Fine, J. P. (2006). Direct parametric inference for the cumulative incidence function. *Applied Statistics* **55**, 187-200
- [24] Jeong, J. H. and Fine, J. P. (2007). Parametric regression on cumulative incidence function. *Biostatistics* **8**, 184-196.
- [25] Jicha, G. A., Abner, E., Schmitt, F. A., Cooper, G. E., Stiles, N., Hamon, R., Carr, S., Smith C. D., Markesbery W. R. (2008). Clinical features of mild cognitive impairment differ in the research and tertiary clinic settings. *Dement Geriatr Cogn Disord* **26**,187-192.
- [26] Joly, P. and Commenges, D. (1999). A penalized likelihood approach for a progressive three-state model with censored and truncated data: Application to AIDS. *Biometrics* **55**, 887C890.
- [27] Joly, P., Commenges, D., Helmer, C., Letenneur, L. (2002). A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* **3**, 433-443.

- [28] Kalbfleisch, J. D., Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863C871.
- [29] Kalbfleisch, J. D. and Prentice, R. L. (2002). The Statistical Analysis of Failure Time Data (2nd). *John Wiley & Sons, Inc.*
- [30] Karvanen, J., Saarela, O., Kuulasmaa, K. (2010). Nonparametric Multiple Imputation of Left Censored Event Times in Analysis of Follow-up Data. *Journal of Data Science* **8**, 151-172.
- [31] Klein, J. P. and Andersen, P. K. (2005). Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics* **61**, 223-9.
- [32] Klein, J. P. (2006). Modelling competing risks in cancer studies. *Statistics in Medicine*. **25**, 1015-34.
- [33] Klein, J.P. (2010). Competing risks. *WIREs Comp Stat* **2**, 333-339.
- [34] Kneib, T., and Hennerfeind, A. (2008). Bayesian semiparametric multi-state models. *Statistical Modeling* **8**, 169-198.
- [35] Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Journal of the Royal Statistical Society, Series C, Applied Statistics* **34**, 201-11.
- [36] Lunn, M. and McNeil, D. (1995). Applying Cox Regression to Competing Risks, *Biometrics* **51**, 524-32.
- [37] Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. (2009). The BUGS project: Evolution, critique, and future directions, *Statistics in Medicine*, **28**, 3049-3067.
- [38] McKhann, G., Drachman, D., Folstein, M., Katzrnan, R., Price, D., and Stadlan, E.M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**,939-944.
- [39] Pan, S.-L., Wu, H.-M., Yen, A., and Chen, T. (2007). A Markov regression random-effects model for remission of functional disability in patients following a first stroke: A Bayesian approach. *Statistics in Medicine* **26**, 5335-5353.
- [40] Pepe, M.S. (1991). Inference for events with dependent risks in multiple end-point studies. *J. Am. Statist. Assoc.* **86**, 770-778.
- [41] Salazar, J.C.(2004). Multi-state Markov models for longitudinal data. Ph.D. Dissertation, Department of Statistics, University of Kentucky.

- [42] Salazar, J.C., Schmitt, F.A., Yu, L., Mendiondo, M. M., Kryscio, R. J. (2007). Shared random effects analysis of multi-state Markov models: Application to a longitudinal study of transitions to dementia. *Statistics in Medicine* **26**, 568-580.
- [43] Schmitt, F.A., Wetherby, M.M.C., Wekstein, D.R., Dearth, C.M.S., Markesbery, W.R. (2001). Brain donation in normal aging: Procedures, motivations, and donor characteristics from the Biologically Resilient Adults in Neurological Studies (BRAiNS) project. *Gerontologist* **41**, 716-722.
- [44] Snowdon, D. A., Greiner, L. H., Mortimer, J. A., Riley, K. P., Greiner, P. A. and Markesbery, W. R. (1997). Brain Infarction and the Clinical Expression of Alzheimer Disease: The Nun Study. *Journal of the American Medical Association* **277**, 813-17.
- [45] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van, der Linde A. (2002). Bayesian Measures of Model Complexity and Fit (with Discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583-616.
- [46] Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 20-22.
- [47] Tsiatis, A., Degruetola, V., Wulfsohn, M. S. (1995). Modeling the relationship of survival longitudinal data measured with error. Application to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, **90**, 27-37.
- [48] Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* **69**, 169-73.
- [49] Wreede, L. C., Fiocco, M., Putter, H. (2010). The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *computer methods and programs in biomedicine* **99**, 261-74;99(3).
- [50] Yu, Binbing, and Ghosh, Pulak (2010). Joint modeling for cognitive trajectory and risk of dementia in the presence of death. *Biometrics* **66**, 294-300.
- [51] Yu, L., Tyas, S. L., Snowdond, D. A., Kryscio, R. J. (2009). Effects of ignoring baseline on modeling transitions from intact cognition to dementia. *Computational Statistics and Data Analysis* **53**, 3334-3343.

## Vita

Yushun Lin

University of Kentucky Department of Statistics

**Date of Birth:** October 8th, 1980

**Place of Birth:** Putian, Fujian Province, China

### Education

---

**Master of Science** in Statistics, University of Kentucky, 2009

**Master of Science** in Physics, University of Nebraska-Lincoln, 2006

**Bachelor of Science** in Materials Science and Engineering, University of Science and Technology of China, 2003

### Employment

---

Research/Teaching Assistant *Aug 2007 to Dec 2011*

Statistics Department, University of Kentucky

Ph. D. Internship (Biometrics) *May 2011 to Aug 2011*

Vertex Pharmaceuticals Inc., Cambridge, MA

Research/Teaching Assistant *Aug 2003 to Jul 2007*

Physics Department, University of Nebraska-Lincoln

### Publications

---

1. Murphy, M. P., Morales, J., Beckett, T. L., Astarita, G., Piomelli, D., Weidner, A., Studzinski, C. M., Dowling, A. L. S., Wang, X. H., LeVine, H., Kryscio, R.

- J., Lin, Y. S., Barrett, E., Head, E. (2011). Changes in Cognition and Amyloid-processing with Long Term Cholesterol Reduction using Atorvastatin in Aged Dogs. *Journal of Alzheimer's Disease* **22**, 135-150.
2. Nelson, P. T., Schmitt, F. A., Lin, Y., Abner, E. L., Jicha, G. A., Patel, E., Thomason, P. C., Neltner, J. H., Smith, C. D., Santacruz, K. S., Sonnen, J. A., Poon, L. W., Gearing, M., Green, R. C.; Woodard, J. L.; Van Eldik, L. J.; Kryscio, R. J. (2011). Hippocampal sclerosis in advanced age: clinical and pathological features. *Brain* **134** , 1506-1518.
  3. Abner, E. L., Kryscio, R. J., Schmitt, F. A., SantaCruz, K. S., Jicha, G. A., Lin, Y., Neltner, J. M., Smith, C. D., Van Eldik, L. J., Nelson, P. T. (2011), "End-stage" Neurofibrillary Tangle Pathology in Preclinical Alzheimer's Disease-Fact or Fiction? *Journal of Alzheimer's Disease* **25**, 445-453.
  4. Sexton, W. M., Lin, Y., Kryscio, R. J., Dawson, D. R. III, Ebersole, J. L.; Miller, C. S. (2011). Salivary Biomarkers of Periodontal Disease in Response to Treatment. *Journal of Clinical Periodontology* **38**, 434-441.
  5. Mirrieles, J., Crofford, L. J., Lin, Y. S., Kryscio, R. J., Dawson, D. R., Ebersole, J. L., Miller, C. S. (2010). Rheumatoid Arthritis and Salivary Biomarkers of Periodontal Disease. *Journal of Clinical Periodontology* **37** , 1068-1074.
  6. Rebel, A., Hassan, Z. U., Boral, L., Lin, Y., Dilorenzo, A., Schell, R. M. (2010). Initial Results of a Structured Rotation in Hematology and Transfusion Medicine for Anesthesiology Residents. *Journal of Clinical Anesthesia* **23** , 469-474.
  7. Sparks, D. L., Kryscio, R. J., Connor, D. J., Sabbagh, M. N., Sparks, L. M., Lin, Y. S., Liebsack, C. (2010). Cholesterol and Cognitive Performance in Normal Controls and the Influence of Elective Statin Use after Conversion to Mild Cognitive Impairment: Results in a Clinical Trial Cohort. *Neurodegenerative Dis* **7** ,183-186.
  8. Dawson, D. R., Wang, C., Danaher, R. J., Lin, Y., Kryscio, R. J., Jacob, R. J., Miller, C. S. (2009). Salivary Levels of Epstein-Barr Virus DNA Correlate with Subgingival Levels, Not Severity of Periodontitis. *Oral Diseases* **15** , 554-559.
  9. Dawson, D. R., Wang, C. M., Danaher, R. J., Lin, Y. S., Kryscio, R. J., Jacob, R. J., Miller, C. S. (2009). Real-Time Polymerase Chain Reaction to Determine the Prevalence and Copy Number of Epstein-Barr Virus and Cytomegalovirus DNA in Subgingival Plaque at Individual Healthy and Periodontal Disease Sites. *Journal of Periodontology* **80**, 1133-1140.