2016

CONTINUOUS TIME MULTI-STATE MODELS FOR INTERVAL CENSORED DATA

Lijie Wan
University of Kentucky, lijiewan0708@gmail.com
Digital Object Identifier: http://dx.doi.org/10.13023/ETD.2016.317

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 needed written permission statement(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) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free 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 an 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 thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Lijie Wan, Student

Dr. Richard J. Kryscio, Major Professor

Dr. Constance Wood, Director of Graduate Studies

CONTINUOUS TIME MULTI-STATE MODELS FOR INTERVAL CENSORED DATA

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
Lijie Wan

Lexington, Kentucky

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

Lexington, Kentucky

2016

ABSTRACT OF DISSERTATION

CONTINUOUS TIME MULTI-STATE MODELS FOR INTERVAL CENSORED DATA

Continuous-time multi-state models are widely used in modeling longitudinal data of disease processes with multiple transient states, yet the analysis is complex when subjects are observed periodically, resulting in interval censored data. Recently, most studies focused on modeling the true disease progression as a discrete time stationary Markov chain, and only a few studies have been carried out regarding non-homogenous multi-state models in the presence of interval-censored data. In this dissertation, several likelihood-based methodologies were proposed to deal with interval censored data in multi-state models.

Firstly, a continuous time version of a homogenous Markov multi-state model with backward transitions was proposed to handle uneven follow-up assessments or skipped visits, resulting in the interval censored data. Simulations were used to compare the performance of the proposed model with the traditional discrete time stationary Markov chain under different types of observation schemes. We applied these two methods to the well-known Nun study, a longitudinal study of 672 participants aged ≥ 75 years at baseline and followed longitudinally with up to ten cognitive assessments per participant.

Secondly, we constructed a non-homogenous Markov model for this type of panel data. The baseline intensity was assumed to be Weibull distributed to accommodate the non-homogenous property. The proportional hazards method was used to incorporate risk factors into the transition intensities. Simulation studies showed that the Weibull assumption does not affect the accuracy of the parameter estimates for the risk factors. We applied our model to data from the BRAiNS study, a longitudinal cohort of 531 subjects each cognitively intact at baseline.

Last, we presented a parametric method of fitting semi-Markov models based on Weibull transition intensities with interval censored cognitive data with death as a competing risk. We relaxed the Markov assumption and took interval censoring into

account by integrating out all possible unobserved transitions. The proposed model also allowed for incorporating time-dependent covariates. We provided a goodness-of-fit assessment for the proposed model by the means of prevalence counts. To illustrate the methods, we applied our model to the BRAiNS study.

KEYWORDS: Longitudinal Data, Multi-State Model, Interval Censoring, Markov, Semi-Markov, NUN Study, BRAiNS Study

Author's Signature: Lijie Wan

Date: July 15, 2016

CONTINOUS TIME MULTI-STATE MODELS FOR INTERVAL CENSORED DATA

By

Lijie Wan

Richard J. Kryscio, PhD
Director of Dissertation

Constance Wood, PhD
Director of Graduate Studies

May 15, 2016
Date

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor Dr. Richard J. Kryscio for his guidance, support and patience during my PhD studies. His wisdom and insights helped me better understand and made possible this work. He encouraged me to not only focus on theoretical research but also on the practical statistical analysis skills, which benefits me greatly both academically and career wise.

Also, I would like to thank my dissertation committee members, Dr. William S. Griffith, Dr. Mai Zhou, Dr. Arnold Stromberg, Dr. Li Chen, and Dr. Erin L. Abner for serving on my Supervisory Committee. I feel grateful for their time and critical suggestions of this dissertation. I also thank Wenjie Lou for valuable comments and discussion on the BRAiNS and Nun project.

In addition to the people I mentioned above, I want to thank all the faculty, students and staffs from the Department of Statistics in University of Kentucky, for all the knowledge and help I received from them.

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

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Overview

In most longitudinal medical studies on progression of healthy individuals to chronic diseases, such as cancer, AIDS, and dementia, the nature of the development is often expressed in terms of distinct health stages, where patients are observed at certain time points and covariate information is collected at several occasions.

Multi-state models (MSM), as generalizations of survival and competing risks models, are the most common models for describing longitudinal failure time data. These models have wide application in modeling the complex evolution of chronic diseases. In epidemiology, multi-state models are used to represent the trajectory of subjects through different discrete states, generally including clinical disease and death.

Handling interval-censored data is considerably more difficult, both analytically and numerically, in MSMs than in survival models and competing risk models, especially for more complex models. The complexity of a MSM mainly depends on the number of states and the possible transitions from these states. The more complex the model, the more difficult it is to define and evaluate the likelihood. For the homogeneous Markov model (HMM), the solution to this problem has long been known, although not widely used in medical research or epidemiology. For non-homogenous Markov Models (NHMM) or semi-Markov models, the problem of inference with interval-censored data is considerably more difficult. One key point is that transition probabilities can be expressed simply in terms of transition intensities in HMM but not in more general multi-state models. Another key point is that that interval-censoring in multi-state models gives rise to a new difficulty, which does not arise in survival models. Generally, several paths

are possible for transitioning from state h to state j between time s and time t, so it is not known which paths occurred [1].

We aim to develop flexible and powerful statistical methods to address the issue of interval-censored data in the application of MSM. In the following, we will introduce the Nun Study and BRAiNS Study, a review of multi-state models, the problems encountered with current methods, and the methodologies we propose to address this problem.

1.2 Background of the Nun Study

The Nun Study is a well-known cohort study designed to assess the influence of early life exposures and cognitive ability on the development of Alzheimer-type dementia and pathology in late life. 672 members of the School Sisters of Notre Dame religious congregation born between 1890 and 1916 and living in the Midwestern, eastern, and southern United States agreed to annual cognitive and functional assessments, and to brain donation [2]. The Nun Study was established at the University of Kentucky in 1991 and moved to the University of Minnesota in 2008. Only the data collected until 2008 is used in this dissertation.

Both time-independent and time-dependent covariates were recorded. Time independent covariates were recorded only once; for example, education level and the gene-related factor, Apolipoprotein E4 carrier status. Time dependent covariates were recorded at each of the follow up assessments.

At each assessment, the cognitive status of each subject was categorized into several different states. In our study, we focused on three cognitive states: Not Serious

Impairment (NSI), Global Impairment (GI) and Dementia. A fourth state, Death, was also included in our model as an important competing risk for states GI and Dementia. The transition flow among these four states is shown in Figure 1.1.

Subject follow-up was planned to last until death. However, some subjects were still alive at the end of the data collection period which results in right censored data. The time to each cognitive state and dementia is subject to interval censoring, due to the fact that each assessment was taken at irregularly spaced discrete time points.

1.3 Background of the BRAiNS Study

The Biologically Resilient Adults in Neurological Studies (BRAiNS) began enrolment in 1989 at the Sanders Brown Center on Aging at the University of Kentucky. The purpose of the BRAiNS project is to study normal aging of the brain in contrast to Alzheimer's disease. Subjects are recruited in phases and receive annual assessments with brain donation at death. All subjects were cognitively intact at study entry.

Using results of annual assessments, a subject was placed into one of several mutually exclusive clinical cognitive states. In our studies, we focused on the transitions between the following states: normal, MCI, dementia and death. Figure 1.2 presents the transition flow and the frequencies for each possible transition.

Right-censored data arises due to subjects' early drop out or the fact that some subjects were still in the normal or MCI state when data collection for the current study ended. Transition time to MCI and dementia are all interval-censored. Cognitive assessments were taken at discrete time points, thus the exact transition times to MCI and Dementia were unknown.

1.4 Multi-State Models

A multi-state model (MSM) is a model for a stochastic process allowing individuals to move among a finite number of states. In biomedical applications, the states might be based on clinical symptoms (e.g. bleeding episodes), biological markers (e.g. CD4 T-lymphocyte cell counts; serum immunoglobulin levels), severity levels of the disease (e.g. stages of cancer or HIV infection) or a non-fatal complication in the course of the illness (e.g. cancer recurrence). A change of state is called a transition, or an event. States can be transient, if the transitions to and from the state are possible, or absorbing, if no transitions can emerge from the state (for example, death) [3-6].

1.4.1 Multi-State Process

Continuous-time multi-state models are based on the theory of a multi-state process, which is assumed to be a stochastic process $\mathrm{X(t)}$ with a finite state space $S = \{1,2,\dots,K\}$. It can be fully characterized by its transition probability matrix or its transition intensity matrix. The transition probability matrix $P(s,t)$ is a $\mathrm{K} \times \mathrm{K}$ matrix, and its (h,j)th entry is

$$p_{hj}(s,t) = P(X(t) = j | X(s) = h, H_{s-}), s < t$$

$P_{hj}(s,t)$ represents the probability of the process being in state j at time t given its state h at time s and the history of the process before time s, H_{s-} .

The transition intensity matrix measures the instantaneous hazard of transition to other states given the current state. The (h,j)th entry of the transition intensity matrix $Q(t)$ at time t has the form:

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} P(X(s + \Delta s) = j | X(s) = h, H_{s-}) / \Delta t , \ if\ h \neq j$$

and

$$\alpha_{hh}(t) = -\sum_{j \neq h} \alpha_{hj}(t).$$

Different model assumptions can be made about the dependence of the transition rates on time. Examples include:

1. Time homogeneous models: the intensities are constant over time *t*.
2. Markov models: the transition intensities only depend on the history of the process through the current state.
3. Semi-Markov models: future evolution not only depends on the current state h, but also on the entry time t_h into state *h*. Therefore, we may consider intensity functions of the general form $\alpha_{hj}(t, t - t_h)$ or, as the special homogeneous case $\alpha_{hj}(t - t_h)$.

1.4.2 Markov Models

The process $(X(t), t \geq 0)$ is Markovian if the transition probabilities and transition intensities are independent of the past history, that is, for any s, t with $0 \leq s < t$, we have

$$P(X(t) = j | X(s) = h, H_{s-}) = P(X(t) = j | X(s) = h)$$

and

$$\alpha_{hj}(t) = \begin{cases} \lim_{\Delta t \to 0} \mathrm{P}(X(t + \Delta t) = j | X(t) = h) / \Delta t & j \neq h \\ -\sum_{k \neq h} \alpha_{hk}(t) & j = h \end{cases}$$

For a Markov process, the future of the process after time t depends only on the state occupied at time t. Under the Markov assumption, the transition probabilities can

be calculated from the intensities by solving the forward Kolmogorov differential equation [4].

1.4.3 Semi-Markov Models

For the semi-Markov model, the transition intensities of the process depend on the time elapsed at the current state. These processes are generalizations of both continuous and discrete parameter Markov processes with countable state spaces. An issue in using a semi-Markov model is identifying the time origin, the exact time of entrance into the initial state.

1.4.4 Modeling Intensities

Covariates in multi-state models are often incorporated through the transition intensity functions to explain differences among individuals in the course of the disease progression. A popular choice is the proportional hazards model, which has the following form

$$\alpha_{hj}(t|\boldsymbol{Z}_i) = \alpha_{hj,0}(t)exp\left(\boldsymbol{\beta}_{hj}^T\boldsymbol{Z}_i\right).$$

In an MSM, the transition intensities define the hazard of a movement from one state to another. These functions can also be used to determine the mean sojourn time in a given state and the number of individuals in different states at a certain moment.

1.5 A Review of Methods for Dealing Interval Censoring Data

Markov models are popular tools for analysis of longitudinal data, since the assumption simplifies statistical modelling. According to this assumption, the transition to the next state only depends on the current state, ignoring any previous history of the process.

Only a few studies have been carried out regarding NHMM in the presence of interval-censored data. Most of the literature is limited to the three-state models. One of the first such studies was that of Hsien, et al.[7], who examined a three-state progressive non-homogenous Markov model with the incorporation of Weibull distribution or the piecewise exponential model to accommodate non-constant transition rates. Hout, et al. [8] extended this approach by including the possibility to move directly from the health state to the death state, resulting in an illness-death model. Both the Weibull distribution and the piecewise-constant model are investigated to deal with the time dependency of the intensities. The model is extended by using logistic regression models for both misclassification probabilities and the latent distribution of the states at baseline. Nonparametric approaches to NHMMs may follow two paths: one is the completely non-parametric approach as a generalization of the Turnbull approach; the other implies a restriction to smooth intensities models. The first explicit non-parametric treatment of interval-censored observations from a MSM in continuous time is given by Frydman, who studied a progressive three-state model [9], and a special case of the illness-death model [10]. The penalized likelihood approach already proposed for interval-censored survival data [11] was extended to a three-state progressive model by Joly and Commenges [12] and to the illness-death model by Joly, et al. [13].

However, in many applications, the Markov assumption might not be appropriate and may lead to biased conclusions. A semi-Markov model would be more appropriate in this case, to allow the transition intensities of the process to depend not only on the current state, but also the time elapsed in the current state.

There has not been much literature on the application of semi-Markov models for interval censored data. Satten [14] proposed non-parametric estimators based on an EM algorithm in the case of a unidirectional model without covariates. Foucher [15, 16] defined a semi-Markov model based on a generalized Weibull hazard function. The model is defined by the probability of transition among states and, independently, the holding time it takes for that transition to occur. The holding times of the underlying process are assumed to follow a generalized Weibull distribution. Kapetanakis [17] recently presented a parametric method of fitting semi-Markov models with piecewise-constant hazards in the presence of left, right, and interval censoring.

Our research is motivated by two longitudinal studies investigating cognitive ability in the older population, the Nun Study and BRAiNS Study. Previous work was carried out by Salazar, et al. [18] , Yu, et al. [19], Abner, et al. [20] and Kryscio, et al. [21, 22]. Salazar, et al. [18] proposed a multi-state Markov model with shared random effects to estimate the one-step transition matrix. In their model, polytomous logistic regression models with shared random effects were first introduced to account for the correlations between observations among the same subjects. The likelihood functions were constructed by integrating the random effect out. Their simulation study showed the approximation to their integral produced reasonable estimates of the unknown model parameters in the one-step transition matrix, and that these parameter estimates are robust across a spectrum of distributions for the shared random effect. However, the model approximated the joint distribution of the response variable using a conditional distribution given the baseline outcome of the response variable, which could produce a so-called “baseline confounding” problem. Yu, et al. [19] extended Salazar's model to

include the information of the baseline state to address this limitation by accommodating the baseline confounding in the Markov model using shared random effects approaches. However, under a shared random effects model, separating the baseline distribution from the overall model likelihood can lead to underestimation of the effects of risk factors on the one-step transitions. Abner, et al. [20] expanded Salazar's model to investigate the transient nature of MCI by including a clinically determined MCI state as an outcome. The multistate Markov chain with three transient states (normal cognition, aMCITB, and mMCITB), one quasi-absorbing state (MCICC), and two absorbing state (death and dementia) was used to model the probability of maintaining the current state or moving to a different state at the next assessment. Here aMCITB and mMCITB represent amnestic and mixed forms of the MCI state as determined by cognitive tests (test based). However, transitions to MCICC and dementia states are still assumed to have occurred on the date of assessment. The model also ignores any transitions among the transient states between regularly scheduled assessments. While these methods are all easy to implement and quite useful, several assumptions need to be satisfied. First, time intervals between two consecutive assessments are required to be equally spaced. However, in many observational longitudinal studies, it is very common to have unequally spaced longitudinal data resulting from uneven assessments or skipped visits. Second, the exact transition times are assumed to occur exactly at the discrete assessment time points since modeling assumptions do not permit the inclusion of interval censoring-type approaches. In reality, interval censored data commonly exist and transitions may take place at any time. Third, those models assume no censored states exist. They assume all possible transitions could be observed between two consecutive assessments. Any transitions

among the transient states during the follow-up assessments are ignored since those models use only the state of the individual at the next assessment. In fact, in most studies we might not be able to tell whether a patient went through other transient states before the following assessment. From the above, the Markov chain model does have some limitations since they rely on a discrete-time model.

Kryscio, et al. [21, 22] applied a semi-Markov model, which is defined by the probability of the transition among states and independently the holding time it takes for that transition to occur. The model is useful to identify risk factors for transitions to MCI and dementia by adjusting the competing risk of death. However, this model is still based on a discrete model, and has the limitation of ignoring interval-censored transition times or unobserved transitions between successive assessments.

1.6 Outline of the Dissertation

The remainder of this dissertation is organized as follows.

In Chapter 2, a continuous time version of homogenous Markov multi-state model with backward transitions is proposed to handle the uneven follow-up assessments or skipped visits, resulting from the interval censored data. Simulations are used to compare the performance of the proposed model with the traditional discrete time stationary Markov chain under different types of observation schemes.

In Chapter 3, we construct a non-homogenous Markov model for this type of panel data. The baseline intensity is assumed to be Weibull distributed to accommodate the non-homogenous property. The proportional hazards method is used to incorporate risk factors into the transition intensities.

In Chapter 4, we present a parametric method of fitting semi-Markov models based on Weibull transition intensities to interval censored cognitive data with death as a competing risk. We relax the Markov assumption and take into account interval censoring by integrating out all possible unobserved transitions. The proposed model also allows for incorporating time-dependent covariates. A goodness-of-fit assessment is provided for the proposed model by the means of prevalence counts.

Finally in Chapter 5, we summarize the work and offer some potential areas for future study.

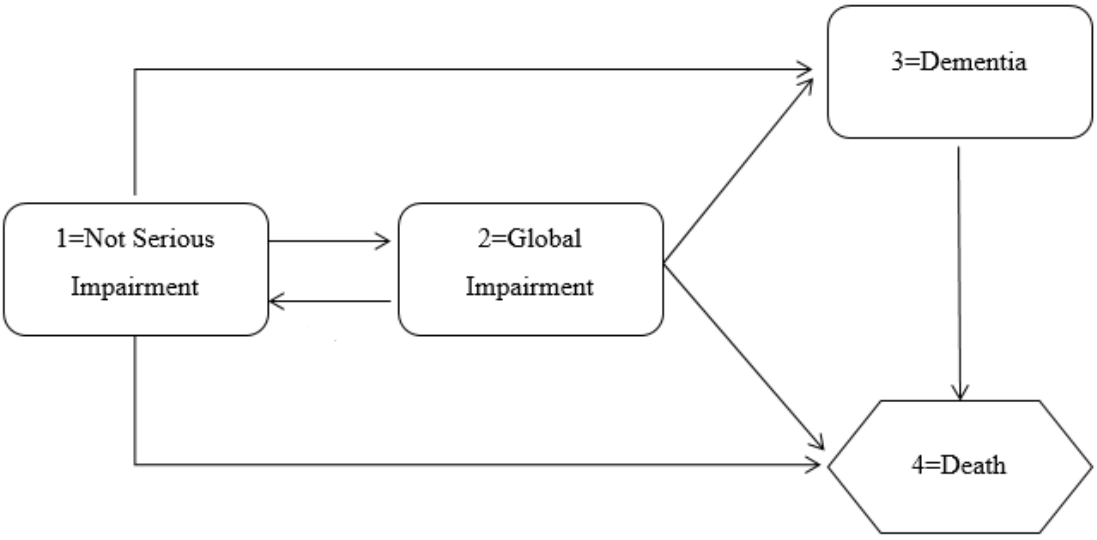

Figure 1.1 Transition flows among the four states recorded in the Nun Study data

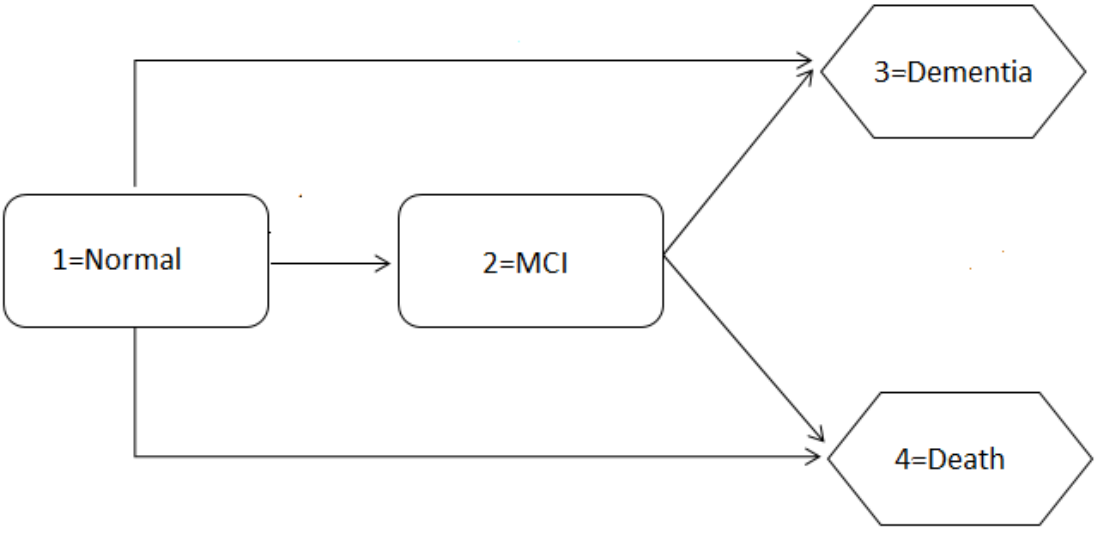

Figure 1.2 Transition flows among the four states recorded in the BRAiNS data

Chapter 2 A Comparison of Discrete-time and Continuous-time Multi-state Models

2.1 Introduction

In most longitudinal medical studies on the progression of healthy individuals to chronic diseases, such as cancer, AIDS, and dementia, the natural development is often expressed in terms of distinct states. The analyses in such studies where individuals may transition among several states are often performed by using multi-state models (MSMs). There are two major types of multi-state models in literature, one is based on discrete-time Markov chain, and the other one is based on continuous-time Markov process. These two types of modeling techniques are related in certain ways, and both enable researchers to study transitions between different disease states simultaneously. However, the two types of models are constructed under different assumptions, and might generate different results and conclusions under certain cases. Thus, researchers need to be careful when deciding which models to use in real data applications.

Multi-state models based on the discrete-time Markov chain have become popular in analyzing longitudinal data collected in chronic disease studies. Such models are also called Markov chain transitional models [23] in the literature. Kryscio, et al. [24] used a Markov chain model to identify risk factors associated with transitions from cognitively normal to various forms of mild cognitive impairment (MCI) and then from MCI into early dementia, with death before dementia as a competing state. A series of polytomous logistic models were used to model the one step transition probabilities, and they focused on the effects of baseline age, education, sex, family history of dementia, and APOE4 status on the transition probabilities.

Use of Continuous-time MSMs has grown quickly in literature. A continuous-time MSM is a model for a continuous time stochastic process allowing individuals to move among a finite number of states [4]. There exists an extensive literature on Continuous-time MSMs [4-6, 25]. Applications of continuous-time MSMs can be found in liver cirrhosis [26], dementia [11-13, 27], etc.

In real data applications, the observation schemes vary among different studies. In some studies, investigators are able to collect the data at equally spaced time points, for example once a month or once a year. In this case, the resulting longitudinal data will be evenly spaced. In other studies, collecting the data at equal time intervals is unrealistic; in these cases the longitudinal data will be unevenly spaced. Both types of MSMs are widely used in applications to model similar longitudinal data without considering the observation schemes. In this manuscript we will conduct a comparison study between the two types of models. To the best of our knowledge, there are few studies in the literature that compares these methods.

The rest of this chapter is structured as follows. In Section 2.2, the discrete-time MSM and continuous-time MSM are introduced respectively. In Section 2.3, a simulation study is conducted to compare the two modeling methods under different observation schemes. Section 2.4 applies the two methods to a real dataset, the Nun study. Conclusion and discussion are provided in Section 2.5.

2.2 Discrete-time and continuous-time multi-state models

For a chronic disease with K possible outcome states, we could write the underlying disease process as $X(t) \in \{1,2, \dots, K\}, t \geq 0$. Here, the value of $X(t)$ denotes the occupied disease state at time t. Suppose an individual has observations at time

points $\boldsymbol{T} = (\ t_0, t_1, \dots, t_{\mathrm{m}})$, we write $\boldsymbol{X} = (X_0, X_1, \dots, X_m)$ the corresponding occupied states such that $X_l = X(t_l), l = 1, 2, \dots, m$. The initial state X_0 is usually given.

2.2.1 Discrete-time multi-state model

In a discrete-time multi-state model, the longitudinal data are modeled through a joint probability mass function $P(X_0, X_1, \dots, X_m\ )$. The observation time points $\boldsymbol{T} = (\ t_0, t_1, \dots, t_{\mathrm{m}})$ are ignored under the assumption that the data are evenly spaced. In most applications, the outcome data $(X_0, X_1, \dots, X_m)$ are assumed to follow a discrete-time Markov chain, in which we have

$$P(X_0, X_1, \dots, X_m\ ) = P(X_0) \times P(X_1|X_0) \times \cdots \times P(X_m|X_{m-1}).$$

The one-step transition probability from state h to state j at lth step can be written as

$$P_{hj,l} = P(X_l = j|X_{l-1} = h).$$

Thus, the joint probability mass function $P(X_0, X_1, \dots, X_m\ )$ can be characterized by the one-step transition probability matrix

$$\boldsymbol{P_l} = \begin{pmatrix} P_{11,l} & \cdots & P_{1K,l} \\ \vdots & \ddots & \vdots \\ P_{K1,l} & \cdots & P_{KK,l} \end{pmatrix}.$$

The rows of $\boldsymbol{P_l}$ satisfy the condition $\sum_{j=1}^{K} P_{hj,l} = 1$. The Markov chain is often assumed to be time homogenous. In this case, we have$\boldsymbol{P_l} = \boldsymbol{P}$ and $P_{hj,l} = P_{hj}$, which is a constant of time.

Baseline covariates $\boldsymbol{Z}$ are usually linked to the transition probabilities through a series of polytomous logistic regressions

$$log\left(\frac{P_{hj}}{P_{hh}}\right) = \beta_{hj,0} + \boldsymbol{\beta_{hj}^T Z}, \qquad j \neq h.$$

There are K possible polytomous logistic regressions, one model for each row of the transition probability matrix. When the model only involves baseline covariates, standard software such as PROC LOGISTIC and PROC CATMOD (SAS Institute, Inc.; Cary NC) [28] can be used to fit each logistic model separately.

2.2.2 Continuous-time multi-state model

In a continuous-time multi-state model, the transition process is modeled as a stochastic process. The longitudinal data are allowed to be unevenly spaced. We can write the transition probability from state h at time s to state j at time t as

$$P_{hj}(s,t|H_{s-}) = P(X(t) = j|X(s) = h, H_{s-}),\ \mathrm{s} < \mathrm{t}.$$

Here, H_{s-} is the history of the process up to time s. For a Markov process, the transition probabilities is independent of the past history before time s. In this case, we have

$$P_{hj}(s,t) = P(X(t) = j|X(s) = h),\ \mathrm{s} < \mathrm{t}.$$

The transition probabilities can be fully characterized by the corresponding transition intensities, which have the following definition

$$\alpha_{hj}(t) = \lim_{\Delta t \to 0} P_{hj}(t + \Delta t, t)/\Delta t,\ j \neq h.$$

Similar to the hazard function in survival models, the transition intensities measure the instantaneous hazard of transition from the current state h to another state j. For $j = h$, we have

$$\alpha_{hh}(t) = -\sum_{j \neq h} \alpha_{hj}(t).$$

Different assumptions can be made about the dependence of the transition intensities on time. In this study, we focus on time homogenous models. In a time homogenous model, we have $\alpha_{hj}(t) = \alpha_{hj}$.

Covariates of interest can be incorporated into the transition intensities using the Cox proportional hazards regression model, which has the following form

$$\alpha_{hj}(\boldsymbol{Z}) = \alpha_{hj,0}\, exp\left(\boldsymbol{\beta_{hj}^T Z}\right) = exp\left(\beta_{hj,0} + \boldsymbol{\beta_{hj}^T Z}\right).$$

Here, $\alpha_{hj,0} = exp\left(\beta_{hj,0}\right)$ is called the baseline intensity from state h to state j.

Write the transition intensity matrix as

$$\boldsymbol{Q} = \begin{pmatrix} \alpha_{11}(\boldsymbol{Z}) & \cdots & \alpha_{1K}(\boldsymbol{Z}) \\ \vdots & \ddots & \vdots \\ \alpha_{K1}(\boldsymbol{Z}) & \cdots & \alpha_{KK}(\boldsymbol{Z}) \end{pmatrix},$$

and write the transition probability matrix as

$$\boldsymbol{P}(s,t) = \begin{pmatrix} P_{11}(s,t) & \cdots & P_{1K}(s,t) \\ \vdots & \ddots & \vdots \\ P_{K1}(s,t) & \cdots & P_{KK}(s,t) \end{pmatrix}.$$

For a time homogenous model, $\boldsymbol{P}(s,t)$ can be calculated in terms of the transition intensity matrix $\boldsymbol{Q}$ using the Kolmogorov differential equation [5]

$$\boldsymbol{P}(s,t) = \boldsymbol{P}(t-s) = Exp((t-s)\boldsymbol{Q})$$

Estimation of the model can be done using the maximum likelihood method. Given an individual has observations at time points ($t_0, t_1, \ldots, t_{\mathrm{m}}$) and corresponding observed states ($X_0, X_1, \ldots, X_m$), its likelihood contribution can be calculated as

$$L = P(X_0, X_1, \ldots, X_m) = P(X_0) \times P_{X_0 X_1}(t_0, t_1) \times \cdots \times P_{X_{m-1} X_m}(t_{m-1}, t_m).$$

Through the transition intensities, we are able to calculate the transition probabilities at any given time period. Thus, we are able to handle unevenly spaced longitudinal data.

We can also handle transitions with exact transition times. Death is an important competing risk in many chronic diseases and is often included in the model. The exact time of death will be recorded, while the state just before death might be unknown. Suppose the last state $X_m = K$ is death and t_{m} is the time of death. In this case, the likelihood contribution can be calculated as

$$L = P(X_0, X_1, \dots, X_m)$$

$$= P(X_0) \times P_{X_0 X_1}(t_0, t_1) \times \cdots \times \left(\sum_{j \neq K} P_{X_{m-1} j}(t_{m-1}, t_m)\, \alpha_{jK}(\boldsymbol{Z}) \right)$$

2.2.3 Relationship between the two models

Two types of models are constructed under different assumptions about the response data. The discrete-time MSM assumes the transitions follow a Markov chain. However, the continuous-time MSM assumes the transitions follow a continuous-time Markov process. Thus, the covariates coefficients in the two types MSMs have different interpretations. The discrete-time MSM incorporates covariates into the model through a series multinomial logit regressions; the corresponding coefficients have the log odds ratio interpretation. The continuous-time MSM incorporates covariates trough transition intensity functions by proportional hazard regressions; the corresponding coefficients have the log hazard ratio interpretation.

The relationship between the two types of models is linked through their one step transition probabilities. Note that in our notation $\boldsymbol{P}$ is the one step transition probability for the discrete-time model and $\boldsymbol{P}(t-s)$ is the transition probability matrix from time s

to time *t* for the continuous-time model. Suppose the time interval between two assessments equals one time unit; thus we have $\boldsymbol{P} = \boldsymbol{P}(1)$.

2.3 Simulation Study

In chronic disease studies, the collected longitudinal data are often not evenly spaced. In this section, we conduct simulation studies to compare the performance of the two types of MSMs under different observation schemes. The comparisons are taken under three types of observed data:

(1) Evenly spaced data: the time intervals between two consecutive observations are all equal to 1 year;

(2) Unevenly spaced data 1: the time intervals between two consecutive observations follow a truncated Normal distribution with mean 1 and standard deviation 0.5, left truncated at 0.25.

(3) Unevenly spaced data 2: the time intervals between two consecutive observations follow a Normal distribution with mean 1 and standard deviation 1.5, left truncated at 0.25.

We focus on the one-year transition probability estimates (P_{hj}). Comparisons are made by their percent biases (% bias) for the two methods under these three types of observed data.

Data are generated from a four-state model with state 1 and state 2 representing two transient states, and state 3 and state 4 representing two absorbing states. The true model has the following transition intensity matrix:

$$\boldsymbol{Q} = \begin{pmatrix} \alpha_{11} & \alpha_{12,0}exp(\beta_{12}Z) & \alpha_{13,0}exp(\beta_{13}Z) & \alpha_{14,0} \\ \alpha_{21,0}exp(\beta_{21}Z) & \alpha_{22} & \alpha_{23,0}exp(\beta_{23}Z) & \alpha_{24,0} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Here, Z is a binary baseline covariate. In our simulation study, Z follows a Bernoulli distribution with probability of 0.4 with value 1. We set the baseline intensities

$$(\alpha_{12,0}, \alpha_{13,0}, \alpha_{14,0}, \alpha_{21,0}, \alpha_{23,0}, \alpha_{24,0}) = (0.25, 0.03, 0.05, 0.2, 0.15, 0.05),$$

and the regression coefficients

$$(\beta_{12}, \beta_{13}, \beta_{21}, \beta_{23}) = (0.5, -0.2, -0.3, 0.15).$$

For all three observation schemes, each subject has up to 30 observations. If a patient is still at state 2 or state 3 after 30 years, it will be right censored at year 30. The exact transition times to state 4 are recorded, while the transition time to state 1, 2, or 3 are all interval censored because of the discrete time observations as we described above.

Simulations are set to 1000 iterations, with each containing 500 subjects. For simplicity, all subjects start at state 1. All calculations are done by using the "msm" package [29] in R and the PROC IML [30] and PROC CATMOD [28] procedures in SAS 9.3 system.

Table 2.1 and Table 2.2 list the percent bias of the one year transition probabilities by discrete-time MSM and by continuous-time MSM respectively. The results show that the discrete-time MSM and continuous-time MSM work equally well when the data is evenly spaced. Since the calculation of transition probabilities through the transition intensities are usually complicated, discrete-time MSM has the computational advantage over the continuous-time MSM.

When the collected longitudinal data are unevenly spaced, the discrete-time MSM will provide biased estimates for the one year transition probabilities. We may observe that the biases of the estimations of one year transition probabilities increase as the spacing gets more uneven. For example, the percent bias of the transition probability estimate from state 1 to state 3 with the covariate $Z = 1$ by the discrete-time Markov MSM could be as large as 20% in unevenly spaced data with relative less the observation time interval variation (unevenly spaced data 1 in the tables), and increase to 69% in unevenly spaced data with relatively larger observation time interval variation (unevenly spaced data 2 in the tables). For the same case, the percent bias of transition probability estimate from state 1 to state 3 with the covariate $Z = 1$ by the continuous-time Markov MSM is only 1.3% in unevenly spaced data 1, and 1.8% in unevenly spaced data 2. Thus, in those longitudinal chronic disease studies in which the actual visit times deviate from the planned visit times, with possible skipped visits, continuous-time MSMs are recommended.

2.4 Application to the Nun Study

In this section, we apply both the discrete-time MSM and continuous-time MSM to the Nun Study dataset. The models include four states: Not Serious Impairment (NSI), Global Impairment (GI), Dementia, and Death. The transition flows and frequencies among these states are shown in Figure 2.1.

A total of 55 subjects were excluded from the study due to missing APOE4 genotype (55 or 8.18%). The final analytic sample used in the study consists of 617 subjects having 3312 observations. At baseline, 440 (71.3%) subjects were in state NSI; 60 (9.7%) subjects were already in state GI and 117 (19.0%) subjects have already

developed dementia. At the end of the study, there were 74 subjects who survived without dementia or censored before converting to dementia, 279 subjects who developed dementia, 264 who died without dementia, and 263 subjects who died with dementia.

Even though the study was designed to conduct cognitive assessments annually, the actual number of total assessments and the time interval between two consecutive assessments varied across subjects. The number of assessments ranges from 2 to 12 with an average of 6 assessments. The time interval between two assessments ranges from 0.01 year to 10 years, with an average of 1.4 ± 0.6 years. Figure 2.2 presents the histogram of the time intervals between two consecutive assessments up to 4 years.

We considered two risk factors in our four-state model: baseline age and APOE4 (1=at least one ε4 allele, 0= noε4 allele). The baseline ages range from 75.37 to 102.01 with mean 83.45 ± 5.53. In the model, baseline ages were centered at age 75. There are 141 (22.85%) subjects with at least one APOE4 allele. Table 2.3 lists the odds ratios of these two risk factors estimated by the discrete-time MSM described in Section 2.1. And Table 2.4 lists the hazard ratios of these two risk factors estimated by the continuous-time MSM described in Section 2.2. Both models show baseline age has significant effects on transitions from NSI to GI, Dementia and Death, from GI to Dementia, and from Dementia to Death; and APOE4 has significant effects on transition from NSI to GI.

The two models differ on the transition probability estimations. Figure 2.3 plots the estimated transition probabilities from NSI to dementia and from GI to dementia for an 80-years-old subjects with and without APOE4. The plots indicate that the discrete-time model has relatively lower long-term transition probabilities from NSI to dementia and from GI to dementia than the continuous-time model.

To see which model fits the data better, we conducted a goodness-of-fit analysis by using prevalence counts [31]. Figure 2.4 presents the comparison of observed prevalence and expected prevalence counts from both discrete-time and continuous-time models. The dot circle line is the observed prevalence counts; the dashed line is the expected prevalence counts estimated from the discrete-time model; and the solid line is the expected prevalence counts estimated from the continuous-time model. In general, the expected prevalence estimated from the continuous-time model is closer to the observed prevalence than the expected prevalence estimated from the discrete-time model. This provides some evidence that the continuous-time model fits the data better than then discrete-time model.

2.5 Discussion

In longitudinal chronic disease studies, the natural development of a chronic disease is often expressed in terms of distinct states and MSMs are widely used to model the progression of individuals through these states. Most studies focused on modeling the true disease progression as a discrete time Markov chain. While Markov chain models can accommodate the simultaneous analysis of multiple events of interest and inclusion of competing risks through the states defined in the model, use of Markov chains have some potential limitations. As it requires the time intervals between two consecutive assessments are all equal among subjects, and it does not allow unobserved transitions between two consecutive assessments. In real studies, the data are often unevenly spaced and multiple unobserved transitions may take place between cycle assessments. A more general model, continuous-time MSM could be an alternative approach which can accommodate the evenly spaced data under different types of observation schemes.

To the best of our knowledge, this research is the first to compare the performance of the widely used discrete-time multi-state model with the continuous-time multi-state model for unevenly spaced data. The simulation study compares the one year transition probability under three types of observed data, one evenly spaced data and two unevenly spaced data. The results show that when the longitudinal observations are evenly spaced, both versions of MSMs work equally well. Since the calculation of transition probabilities through the transition intensities is usually complicated, the discrete-time MSMs have the computational advantage over the continuous-time version MSMs. When longitudinal observations are unevenly spaced, the discrete-time MSMs would be biased. In this case, the continuous-time MSMs are recommended.

In the application of the Nun's data, the discrete-time model has relative worse performance compared to the continuous-time model. Both models provided similar results of the effects of baseline age and APOE4 in the model. However, the estimations of the transition probabilities are different by the two models. The discrete-time model has relative lower long-term transition probability estimations from state NSI to dementia and from state GI to dementia. The average time interval between two consecutive assessments was 1.4±0.6 years (larger than 1 year assumption of the discrete-time model) in the Nun's data, which is one of the reason the discrete-time model underestimates the long-term transition probabilities from NSI to dementia and from GI to dementia.

In conclusion, discrete-time Markov chain models are useful tools for survival analysis that allow for more nuanced modeling that is available in most standard time to event methods. However, most journal readers and reviewers may readily comprehend the results from discrete-time Markov chain models, but they may lack familiarity with

the underlying statistical assumptions. If so, they may neglect to challenge investigators to demonstrate these assumptions are tenable [32]. A continuous time MSM could be an alternative approach and should have a potential to being used much more by practitioners, although the lack of knowledge of the available software may be responsible for its lack of popularity. Given that improper use of Markov models may result in biased estimation, perhaps some standardization in the reporting of MSM results and assumption verification is needed.

Table 2.1 Percent bias of one year transition probability for each path by the discrete-time multi-state model under three observation schemes.

Transition	Evenly Spaced Data		Unevenly Spaced Data 1		Unevenly Spaced Data 2	
	Z=0	Z=1	Z=0	Z=1	Z=0	Z=1
1 to 1	-0.29%	0.22%	0.77%	2.50%	-2.00%	-1.10%
1 to 2	0.40%	-0.42%	-2.00%	-10.00%	-5.50%	-14.00%
1 to 3	-0.74%	1.00%	-1.40%	20.00%	38.00%	69.00%
1 to 4	-0.38%	0.76%	-3.30%	7.00%	19.00%	28.00%
2 to 1	0.27%	-0.39%	-3.10%	-7.90%	-7.20%	-12.00%
2 to 2	-0.18%	0.21%	-0.51%	2.90%	-3.30%	-0.09%
2 to 3	0.45%	-0.45%	3.60%	-5.90%	16.00%	6.60%
2 to 4	0.66%	-0.94%	7.20%	-9.50%	27.00%	5.20%

Table 2.2 Percent bias of one year transition probability for each path by the continuous-time multi-state model under three observation schemes.

Transition	Evenly Spaced Data		Unevenly Spaced Data 1		Unevenly Spaced Data 2	
	Z=0	Z=1	Z=0	Z=1	Z=0	Z=1
1 to 1	-0.06%	-0.31%	-0.01%	0.12%	0.02%	-0.49%
1 to 2	0.42%	0.51%	-0.04%	-0.45%	-0.06%	0.73%
1 to 3	-0.44%	1.50%	0.86%	1.30%	-0.91%	1.80%
1 to 4	-0.23%	-0.31%	-0.39%	-0.37%	0.64%	0.54%
2 to 1	0.41%	0.23%	0.55%	0.82%	0.74%	-0.14%
2 to 2	-0.08%	0.03%	-0.27%	-0.17%	-0.27%	0.11%
2 to 3	0.22%	-0.08%	0.74%	0.17%	0.62%	-0.48%
2 to 4	-0.72%	-0.72%	0.24%	0.30%	0.10%	0.15%

Table 2.3 Discrete-time MSM results on the Nun's data

Covariates	Transition Path	Coefficient	Std.Err	P value
Intercept	NSI to GI	-2.9425	0.1512	<.01
	NSI to Dementia	-3.9931	0.2213	<.01
	NSI to Death	-3.0251	0.1613	<.01
	GI to NSI	-1.1615	0.3001	<.01
	GI to Dementia	-1.4880	0.2901	<.01
	GI to Death	-0.6328	0.2429	<.01
	Dementia to Death	-1.3189	0.1677	<.01
Baseline Age	NSI to GI	0.1044	0.0171	<.01
	NSI to Dementia	0.1405	0.0225	<.01
	NSI to Death	0.0963	0.0186	<.01
	GI to NSI	0.0101	0.0294	0.73
	GI to Dementia	0.0525	0.0254	0.04
	GI to Death	0.0161	0.0235	0.49
	Dementia to Death	0.0632	0.0137	<.01
APOE4	NSI to GI	0.5799	0.1940	<.01
	NSI to Dementia	0.5064	0.2831	0.07
	NSI to Death	0.3976	0.2191	0.07
	GI to NSI	-0.6509	0.4233	0.12
	GI to Dementia	0.5268	0.3066	0.09
	GI to Death	-0.2533	0.3018	0.40
	Dementia to Death	0.0417	0.1661	0.80

Table 2.4 Continuous-time MSM results on the Nun's data

Covariates	Transition Path	Coefficient	Std.Err	P value
Intercept	NSI to GI	-2.8417	0.1515	<.01
	NSI to Dementia	-4.6231	0.3545	<.01
	NSI to Death	-3.8646	0.2856	<.01
	GI to NSI	-1.7183	0.2711	<.01
	GI to Dementia	-1.8899	0.2428	<.01
	GI to Death	-1.2368	0.2364	<.01
	Dementia to Death	-1.6425	0.1382	<.01
Baseline Age	NSI to GI	0.0882	0.0168	<.01
	NSI to Dementia	0.1277	0.0360	<.01
	NSI to Death	0.0646	0.0321	0.04
	GI to NSI	0.0100	0.0268	0.71
	GI to Dementia	0.0375	0.0183	0.04
	GI to Death	-0.0215	0.0246	0.38
	Dementia to Death	0.0364	0.0106	<.01
APOE4	NSI to GI	0.4447	0.1968	0.02
	NSI to Dementia	0.0403	0.7244	0.96
	NSI to Death	0.3561	0.3711	0.34
	GI to NSI	-0.6053	0.3994	0.13
	GI to Dementia	0.4952	0.2611	0.06
	GI to Death	-0.5631	0.3679	0.13
	Dementia to Death	0.0112	0.1331	0.93

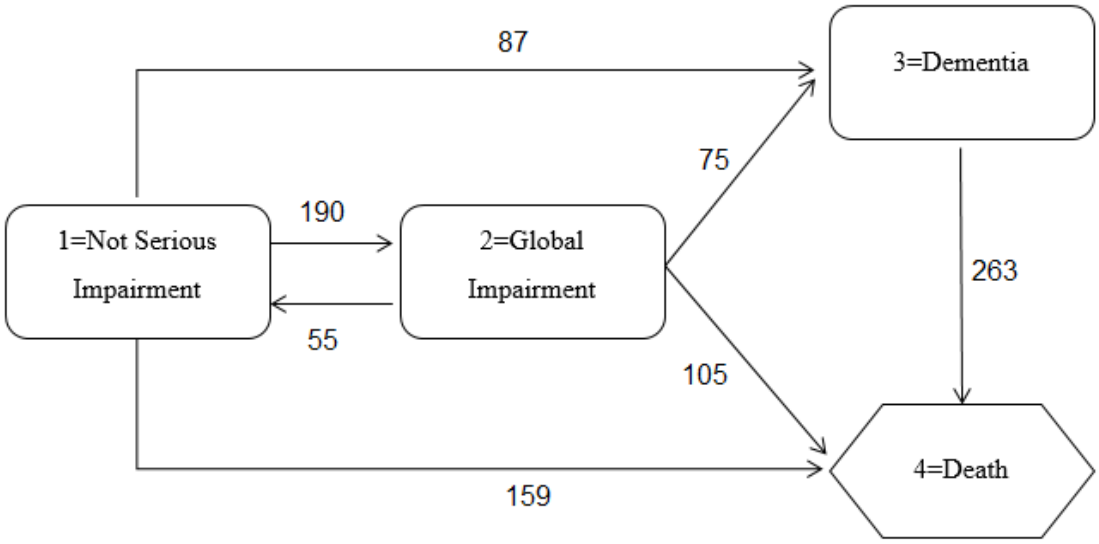

Figure 2.1 Transition flows among the four states recorded in the Nun's data

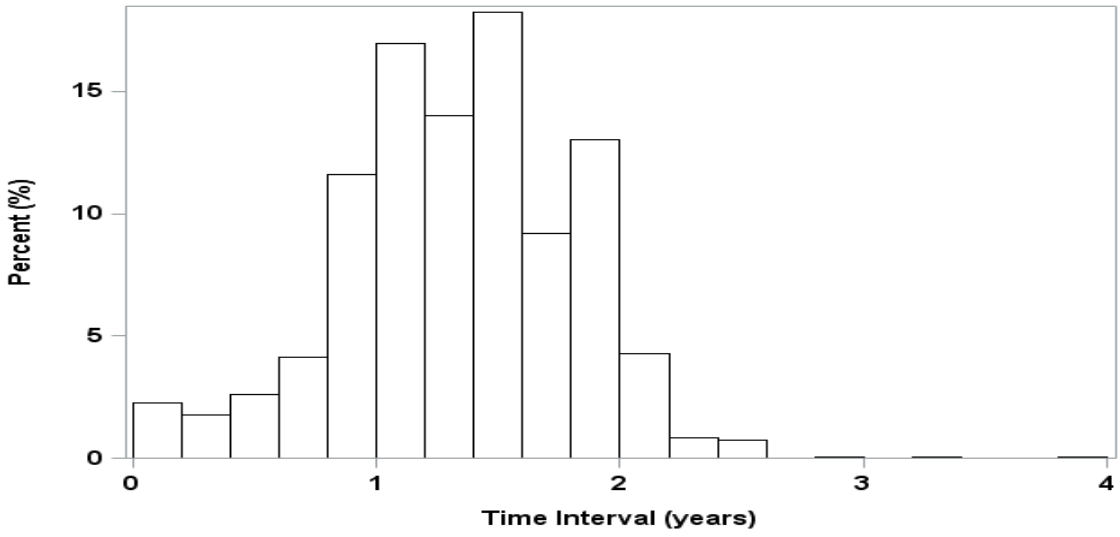

Figure 2.2 Histogram of time intervals between two consecutive assessments.

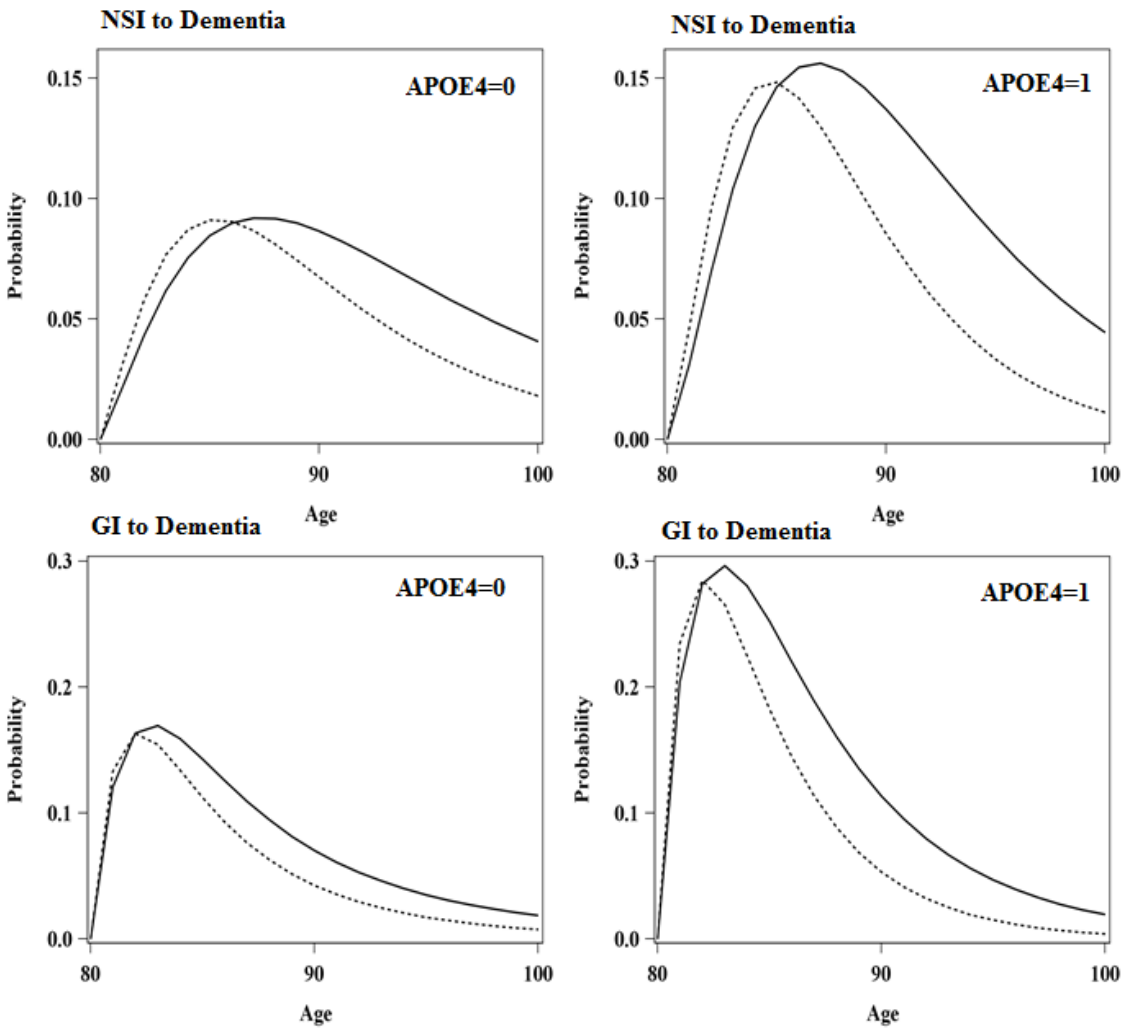

Figure 2.3 Transition probabilities from NSI to dementia and from GI to dementia for an 80 years old subject with and without APOE4.

(Solid line: Transition probabilities estimated by the continuous-time model; and dashed line: Transition probabilities estimated by the discrete-time model.)

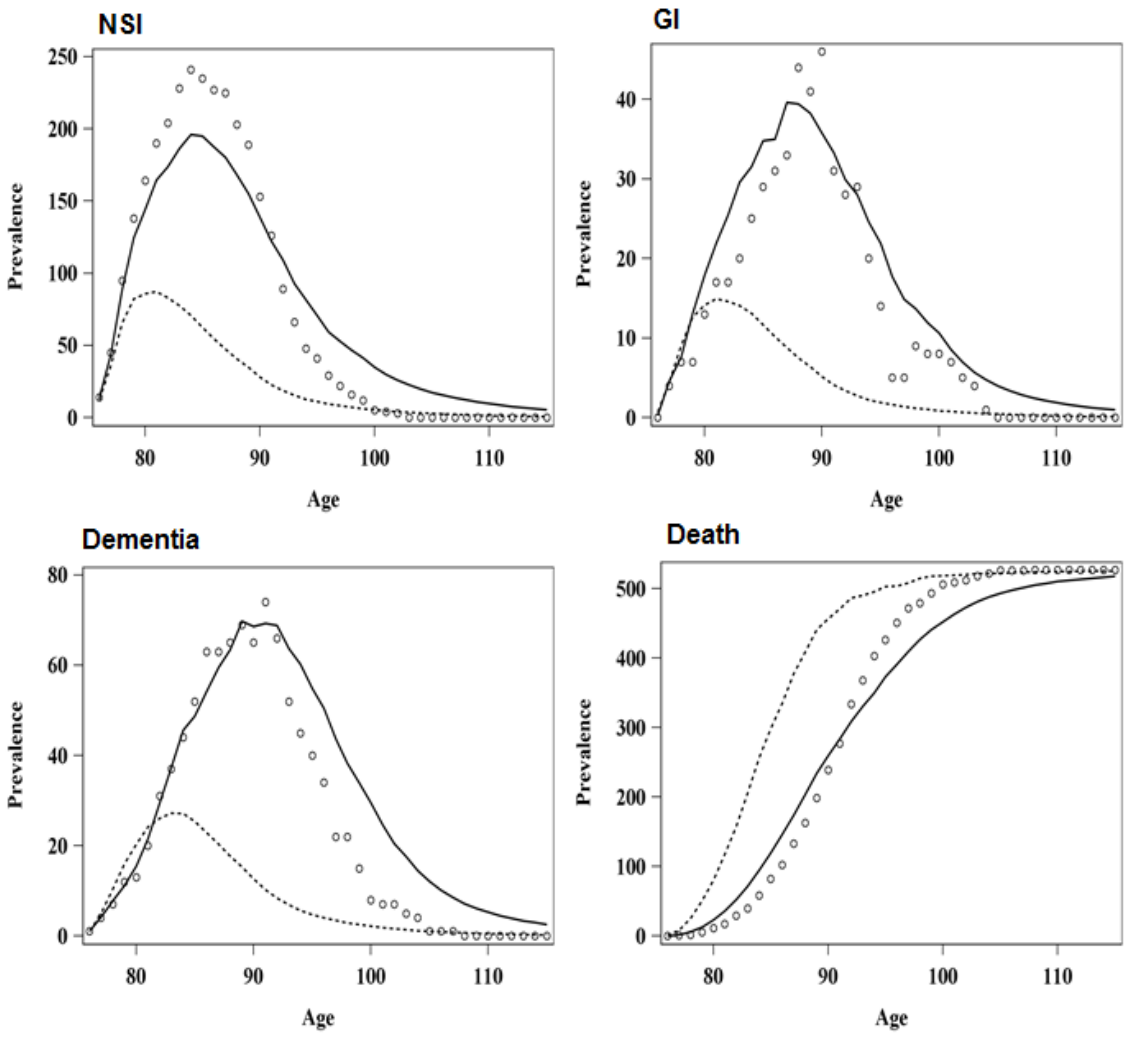

Figure 2.4 Comparison of observed and expected prevalence of the two types of MSMs. (Dot: observed prevalence; solid line: expected prevalence estimated from the continuous-time model; and dashed line: expected prevalence estimated from the discrete-time model.)

Chapter 3 A Non-homogenous Markov Multi-State Model for interval censored transient cognitive states with competing risk

3.1 Introduction

Multi-state interval-censored data are usually handled by time homogenous Markov models (HMM) [33, 34] or piece-wise homogenous models [29]. However, the assumption of time homogeneity would be inappropriate if the disease process is heavily dependent on the time scale considered in the model. In non-homogenous Markov Models (NHMM), the problem of inference with interval-censored data is considerably more difficult. Transition probabilities can be expressed simply in terms of transition intensities in a HMM but not in a more general NHMM.

Only a few studies have been carried out regarding NHMM in the presence of interval-censored data. One of the first such studies was that of Hsien, et al.[35]. They presented a three-state progressive NHMM with the incorporation of Weibull distribution and the piecewise exponential model to accommodate non-constant transition rates. Hout, et al. [8] extended this approach by including the possibility to move directly from the health state to the death state, namely an “illness-death” model. Hubbard and Zhou [27] proposed a non-homogenous four-state model with one absorbing state (death) by using time-transformation. The non-homogenous model is converted to a homogenous model by transforming the time scale by a specific select transformation function. Selecting the appropriate transformation function is the key in their model. However, they did not provide a procedure for selecting the appropriate transformation function, and it is dependent on researcher’s personal judgment to choose which transformation function to use.

In this chapter, we develop a four-state NHMM that allows for interval-censored data as well as the possible unobserved transitions caused by discrete-time observations. The research is motivated by the Biologically Resilient Adults in Neurological Studies (BRAiNS). BRAiNS is a longitudinal study investigating cognitive ability in the older population. We aim to identify and evaluate the effects of the risk factors on the transition among different cognitive states.

The rest of this chapter is structured as follows. In Section 3.2, we describe the data set which motivated this research. In Section 3.3, the four state continuous time Markov model with Weibull assumption is defined. In Section 3.4, a simulation study is conducted to check whether the Weibull assumption is robust. Section 3.5 applies this method to the BRAiNS data. At the last section, we discuss the proposed method and lay down some possible future directions.

3.2 Data

The BRAiNS is a longitudinal cohort of 1,030 older participants at the University of Kentucky's Alzheimer's Disease Center (UK ADC) [36]. Participants consent to extensive annual cognitive and clinical examinations as well as brain donation upon death. Subjects included in the current study (n=531) were assessed at least two times and all subjects were cognitively intact at study entry.

Annual cognitive assessments are administered to each participant and used to classify them into one of three cognitive states: normal, clinical MCI, or dementia. The diagnosis of clinical MCI is based on a consensus team review by the examining physician, neuropsychologist, and the clinical research assistant who administered the

cognitive assessment. A dementia classification results from a clinical consensus diagnosis of dementia.

Mortality has been shown to be an important competing risk for MCI and dementia [22]. Thus, we would also include the state death into our model. Participants were evaluated cognitively and during follow up may die or transition to clinical MCI or dementia. All transitions are unidirectional since it is assumed that once a participant meets the criteria for a diagnosis of clinical MCI (or dementia) he/she does not return to the normal state. In the application to these participants from the BRAiNS cohort, 19 subjects made an apparent reverse transition from clinical MCI to normal, but as discussed in Abner, et al.[20], these were determined to be the result of either underlying medical comorbidities that influenced cognition or diagnostic misclassification; the errant diagnoses were recoded. The exact times of transitions into MCI or dementia are interval censored because of the irregularity of the observation process, while the time of entry into the study and the time of death are known exactly.

3.3 Methodology

3.3.1 The non-homogeneous Markov multi-state model

We consider a four-state model with two transient states and two absorbing states. State 1, normal cognition, and State 2, mild cognitive impairment (MCI), are transient states. State 3, dementia, and State 4, death without dementia, are two absorbing states. See *Figure 3.1* for the transition diagram.

Let $P_{hj}(s,t)$ be the transition probability from state h at age s to state j at age t $(t > s)$. Let $\alpha_{hj}(t)$ be the transition intensity from state h to state j at age t.

Denote $\boldsymbol{Q}(t)$ the transition intensity matrix with the $(h,\ j)$th being $\alpha_{hj}(t)$. Since the model is irreversible, and State 3 and State 4 are absorbing states, we have

$$\boldsymbol{Q}(t) = \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) & \alpha_{13}(t) & \alpha_{14}(t) \\ 0 & \alpha_{22}(t) & \alpha_{23}(t) & \alpha_{24}(t) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Here,

$$\alpha_{11}(t) = -\big(\alpha_{12}(t) + \alpha_{13}(t) + \alpha_{14}(t)\big)$$

and

$$\alpha_{22}(t) = -\big(\alpha_{23}(t) + \alpha_{24}(t)\big).$$

Since age is the major risk factor of MCI, dementia and death [2, 22], and we are mostly interested in age-specific incidence and age-specific mortality, we consider a non-homogeneous Markov model where the intensities depend on age, and chose the actual age of participants as the time scale, t, in the form of transition intensities, rather than the calendar year or years since enrolment.

3.3.2 Proportional hazard regression with Weibull baseline

Other risk factors besides age can also be added to the model through proportional hazards regressions, which has the following form

$$\alpha_{hj}(t|\boldsymbol{Z}) = \alpha_{hj,0}(t - A_0)\, exp\big(\boldsymbol{\beta}_{\boldsymbol{hj}}^{\boldsymbol{T}}\boldsymbol{Z}\big), \qquad t \geq A_0$$

Here, $\boldsymbol{Z}$ is a vector of covariates, such as gender, education level, diabetes, smoking, etc., $\alpha_{hj,0}(t)$ is the baseline intensity for transition from state h to state j, A_0 is the start time of the process, and the covariates coefficients $\boldsymbol{\beta}_{\boldsymbol{hj}}$ are transition specific; in other words, the coefficients for the same covariates on different transition paths are specific to those paths.

We assume the baseline intensities follow the Weibull hazard form with scale parameter $\lambda_{hj} = exp(\beta_{hj,0})$ and shape parameter κ_{hj} to accommodate the non-homogeneous property. We have

$$\alpha_{hj0}(t) = \lambda_{hj}\kappa_{hj}t^{\kappa_{hj}-1} = \kappa_{hj}t^{\kappa_{hj}-1}exp(\beta_{hj,0}), \qquad h = 1,2;\ j = h+1, \dots, 4.$$

Using the Weibull baseline hazards enable us to model a variety shapes of intensity forms. For example, the baseline intensity increases with age when $\kappa_{hj} > 1$; the baseline intensity deceases with age when $0 < \kappa_{hj} < 1$; and the baseline intensity is time homogeneous when $\kappa_{hj} = 1$.

3.3.3 Observation Schemes

Because the study design and the way the cognitive states were determined, the data is left truncated, right censored and interval-censored. First, the data is left truncated. Subjects included in the study are all at normal cognition state at their baseline, and we excluded these subjects who were already in MCI or dementia state from enrolling. Thus the data is left truncated [1]. Second, the data are also right censored. The right censoring occurs when participants drop out of the study before they develop dementia or die, or remain normal cognition or MCI at the end of the study. Since the cognitive states of participants are only assessed at discrete time points, the exact transition time into state MCI and dementia is unknown. The transition times are only known between two consecutive assessed time points where an transition were observed, thus the data is also interval-censored.

The discrete-time observation scheme not only caused interval-censoring, but also lead to unobserved transitions. For example, a participant who is diagnosed with

dementia from normal cognition directly, it is unknown whether the participant has made the transition into MCI first or not. Similarly, if a subject dies with normal cognition at the latest assessment, it is not known whether the subject has made a transition into MCI in the interval between those events. Thus, it is possible that some transitions might not be observed and recorded in the data.

Although the cognition assessments are made at discrete time points, the exact time of death can be retrieved and is recorded in the data. Since there is no cognition assessment at the time of death, the cognitive state just before death is unknown.

There are total 6 possible observed transition paths for a subject, as shown in Figure 3.2. In the BRAiNS data, we set the start time of the process $A_{i,0} = 60$ for all participants since they were all at least 60 years old at their baseline. Also let $A_{i,b}$ be the age at baseline for participant i; $A_{i,1N}$ be the age at the last time participant i is observed in state 1 (normal); $A_{i,20}$ be the age at the first time the participant is observed in state 2 (MCI); $A_{i,2N}$ be the age at the last time the participant is observed in state 2; $A_{i,N}$ be the age at the last time participant i had an observation. We also write $U_{i,hj}$ to be the age at the time participant i transitions from state h to state j, for example, $U_{i,12}$ is the age at the time participant transitions from state 1 to state 2 and $U_{i,23}$ is the age at the time participant transitions from state 2 to state 3. All of $U_{i,12}, U_{i,13}$ and $U_{i,23}$ are interval censored. The transition times to death, which are $U_{i,14}$ and $U_{i,24}$, are known exactly. In our case we have $U_{i,14} = A_{i,N}$ or $U_{i,24} = A_{i,N}$ if the last state recorded is death.

3.3.4 Likelihood

Before we construct the likelihood for the model, we denote two transition probabilities that will help us write the likelihood function. First, we have

$$P_{11}(A_{i,b}, t) = exp\left(-\left(\Lambda_{i,1}(t) - \Lambda_{i,1}(A_{i,b})\right)\right).$$

Here, $\Lambda_{i,1}(t)$ is the cumulative hazard function of subject i for leaving state 1 and it has the form:

$$\Lambda_{i,1}(t) = \int_{A_{i,0}=60}^{t} \left(\alpha_{12}(u) + \alpha_{13}(u) + \alpha_{14}(u)\right)du$$

$$= (t-60)^{\kappa_{12}}\, exp\left(\beta_{12,0} + \boldsymbol{\beta}_{12}^{T}\boldsymbol{Z}\right) + (t-60)^{\kappa_{13}}\, exp\left(\beta_{13,0} + \boldsymbol{\beta}_{13}^{T}\boldsymbol{Z}\right)$$

$$+ (t-60)^{\kappa_{14}}\, exp\left(\beta_{14,0} + \boldsymbol{\beta}_{14}^{T}\boldsymbol{Z}\right)$$

We also have

$$P_{22}(s, t) = exp\left(-\left(\Lambda_{i,2}(t) - \Lambda_{i,2}(s)\right)\right).$$

Here,

$$\Lambda_{i,2}(t) = \int_{A_{i,0}=60}^{t} \left(\alpha_{23}(u) + \alpha_{24}(u)\right)$$

$$= (t-60)^{\kappa_{23}}\, exp\left(\beta_{23,0} + \boldsymbol{\beta}_{23}^{T}\boldsymbol{Z}\right) + (t-60)^{\kappa_{24}}\, exp\left(\beta_{24,0} + \boldsymbol{\beta}_{24}^{T}\boldsymbol{Z}\right)$$

We will discuss the likelihood construction by each of the six paths as shown in Figure 3.2 .

Path (1): the participant has no transition during the study and is stay in state 1 (normal) at the end of study. In this case, the likelihood contribution for this participant would be

$$L_{i,1} = P_{11}(A_{i,b}, A_{i,N}) = exp\left(-\left(\Lambda_{i,1}(A_{i,N}) - \Lambda_{i,1}(A_{i,b})\right)\right).$$

Path (2): the participant has one observed transition from state normal to state MCI and stays in state MCI at the end of study. In this case, we have

$$\begin{aligned} L_{i,2} &= P\left(X(A_{i,1N}) = 1, X(A_{i,20}) = 2, X(A_{i,N}) = 2 \mid X(A_{i,b}) = 1\right) \\ &= \int_{A_{i,1N}}^{A_{i,20}} P\left(X(U_{i,12}-) = 1 | X(A_{i,b}) = 1\right) \alpha_{12}(U_{i,12}) P\left(X(A_{i,N}) = 2 | X(U_{i,12}) = 2\right) dU_{i,12} \\ &= \int_{A_{i,1N}}^{A_{i,20}} P_{11}(A_{i,b}, U_{i,12}) \alpha_{12}(U_{i,12}) P_{22}(U_{i,12}, A_{i,N})\, dU_{i,12} \end{aligned}$$

Path (3): the patient has one observed transition from normal to state dementia. In this case, because of the interval-censoring, there could be two possible true paths; we need to take into account all the information available in such cases of incomplete data. Scenario 1, the subject might have one transition from state normal directly to state dementia at time $U_{i,13}$. Scenario 2, the subject might have two transitions, first transition from state normal to state MCI at time $U_{i,12}$ then transition from state MCI to state dementia at time $U_{i,23}$. Thus, the likelihood of this path has two parts:

$$\begin{aligned} L_{i,3} &= P\left(X(A_{i,1N}) = 1, X(A_{i,N}) = 3 | X(A_{i,b}) = 1\right) \\ &= \int_{A_{i,1N}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,13}) \alpha_{13}(U_{i,13})\, dU_{i,13} \\ &\quad + \int_{A_{i,1N}}^{A_{i,N}} \int_{U_{i,12}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,12}) \alpha_{12}(U_{i,12}) p_{22}(U_{i,12}, U_{i,23}) \alpha_{23}(U_{i,23}) dU_{i,12} dU_{i,23} \end{aligned}$$

Path (4): subject i has one observed transition from state normal to death without dementia. Similar in Path (3), there could be two possible scenarios. The subject might

have just one transition from state 1 directly to state 4 at time $A_{i,N}$; or it might have two transitions, first from state normal to state MCI at time $U_{i,12}$ then transition from state MCI to death at time $A_{i,N}$. Note that in the BRAiNS data the exact age of death is recorded but the state just before death is unknown except dementia. Thus the subject might be at either state normal or state MCI before death. The likelihood for this path can be calculated as follows:

$$\begin{aligned} L_{i,4} &= P\big(X(A_{i,1N}) = 1, X(A_{i,N}) = 4, U_{i,14} = A_{i,N} | X(A_{i,b}) = 1\big) \\ &= p_{11}(A_{i,b}, U_{i,14})\alpha_{14}(U_{i,14}) \\ &\quad + \int_{A_{i,1N}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,12})\alpha_{12}(U_{i,12})p_{22}(U_{i,12}, A_{i,N})\alpha_{24}(U_{i,24}|U_{i,12})dU_{i,12} \end{aligned}$$

Path (5): subject i has two observed transitions, first transition from state normal to state MCI at time $U_{i,12}$ $\left(A_{i,1N} < U_{i,12} \leq A_{i,20}\right)$ and then from state MCI to state dementia at time $U_{i,23}$ $\left(A_{i,2N} < U_{i,23} \leq A_{i,N}\right)$. In this case we have

$$\begin{aligned} L_{i,5} &= P\big(X(A_{i,1N}) = 1, X(A_{i,20}) = 2, X(A_{i,2N}) = 2, X(A_{i,N}) = 3 \,|X(A_{i,b}) = 1\big) \\ &= \int_{A_{i,1N}}^{A_{i,20}} \left(\int_{A_{i,2N}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,12})\alpha_{12}(U_{i,12})p_{22}(U_{i,12}, U_{i,23})\alpha_{23}(U_{i,23})dU_{i,23} \right) dU_{i,12} \end{aligned}$$

Path (6): subject i has two observed transitions, the first is from state normal to state MCI at time $U_{i,12}$ $\left(A_{i,1N} < U_{i,12} \leq A_{i,20}\right)$ and the second is from state MCI to state death at time $A_{i,N}$. In this case we have

$$\begin{aligned} L_{i,6} &= P\big(X(A_{i,1N}) = 1, X(A_{i,20}) = 2, X(A_{i,N} -) = 2\,, X(A_{i,N}) = 4 | X(A_{i,b}) = 1\big) \\ &= \int_{A_{i,1N}}^{A_{i,20}} p_{11}(A_{i,b}, U_{i,12})\alpha_{12}(U_{i,12})p_{22}(U_{i,12}, A_{i,N})\alpha_{24}(A_{i,N})\, dU_{i,12} \end{aligned}$$

3.3.5 Parameter estimation

Since there are no closed forms for likelihood, the trapezoidal rule is used to approximate the integrals to calculate the transition probabilities in terms of transition intensities. The parameter estimation is implemented by maximizing the conditional log-likelihood. In particular, all the calculations are conducted in SAS PROC IML procedure [30]. The log likelihood function can be maximized by the Newton-Raphson Method. The Hessian matrix of the log likelihood function could be approximated by the finite-differences method, and its inverse yields the estimated covariance matrix of the parameters.

3.4 Simulation Study

The main purpose of the simulation study is to examine the sensitivity of the MLEs of the beta estimates in Equation (1) and (2) to the violation of the Weibull assumption on the baseline transition intensities. The goal is to quantify how the different true underlying baseline transition intensities affects the covariate coefficient estimates using the Weibull baseline intensities in the model. The criteria are bias and mean square errors of the MLEs of the covariate coefficients.

Data are generated from a model with the following transition intensity matrix

$$\boldsymbol{Q}(t) = \begin{pmatrix} \alpha_{11}(t) & \alpha_{12,0}(t)exp(\beta_{12}Z) & \alpha_{13,0}(t)exp(\beta_{13}Z) & \alpha_{14,0}(t)exp(\beta_{14}Z) \\ 0 & \alpha_{22}(t) & exp(\beta_{23,0}) & exp(\beta_{24,0}) \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Here, Z is a binary covariate distributed as Bernoulli (0.5) and $(\beta_{12}, \beta_{13}, \beta_{14})$ are the corresponding coefficients. We set the true values of the coefficients$(\beta_{12}, \beta_{13}, \beta_{14}) =$

$(2.0, 1.5, 0)$. For the baseline intensities $\alpha_{12,0}(t), \alpha_{13,0}(t)$, and $\alpha_{14,0}(t)$, we consider three different forms:

(1) Exponential: $\alpha_{1j,0}(t) = exp(\beta_{1j,0}^{E}), j = 2, 3, 4$. We $\left(\beta_{12,0}^{\mathrm{E}}, \beta_{13,0}^{\mathrm{E}}, \beta_{14,0}^{\mathrm{E}}\right) = (-4.2, -4.3, -4)$ for the true model.

(2) Weibull: $\alpha_{1j,0}(t) = \lambda_{1j}\kappa_{1j}t^{\kappa_{1j}} = exp(\beta_{1j,0}^{W})\kappa_{1j}t^{\kappa_{1j}-1}, j = 2, 3, 4$. We set $\left(\beta_{12,0}^{\mathrm{W}}, \beta_{13,0}^{\mathrm{W}}, \beta_{14,0}^{\mathrm{W}}\right) = (-6.5, -6.7, -6)$ and $(\kappa_{12}, \kappa_{13}, \kappa_{14}) = (1.9, 2, 2.1)$.

(3) Gompertz: $\alpha_{1j,0}(t) = \delta_{1j}\, exp\left(\gamma_{1j}t\right) = ex\, p\left(\beta_{1j,0}^{G} + \gamma_{1j}t\right),\ j = 2, 3, 4$. Here, $\left(\beta_{12,0}^{\mathrm{G}}, \beta_{13,0}^{\mathrm{G}}, \beta_{14,0}^{\mathrm{G}}\right) = (-8.3, -8.6, -8.1)$ and $(\gamma_{12}, \gamma_{13}, \gamma_{14}) = (0.2, 0.22, 0.26)$.

To simplify, we set the transition intensities from MCI to dementia and death to be time-homogenous. We set $\left(\beta_{23,0}, \beta_{24,0}\right) = (-1.5, -2.0)$ in the true model. Choice of the model parameters is made to come as close to those estimated from the real dataset of the next section without producing simulations that lead to non-estimable parameters, i.e. the likelihood function fails to converge.

The processes are annually observed starting from State 1 (Normal cognition), with up to 25 follow-up waves. Latent failure time method [37, 38] is used to simulate the multi-state data. Simulations were set to have 1000 iterations, with each containing either 300 or 500 subjects. All simulations are done using the IML procedure [30] in SAS system. The results are presented in Table 3.1 Simulation results of covariate effects for sample sizes 300 and 500.

As expected, increasing the sample size improves the estimates in terms of reducing mean squared error (MSE) and increasing 95% confidence coverage rates. The biases are reduced considerably when the sample size is increased.

The simulation results also show that the proposed model using Weibull baseline intensities provides good estimates of the covariate coefficients even in cases where the true baseline intensities are not Weibull. For all the three intensity forms considered in this study, the bias and mean square error (MSE) of the estimated effects are all relatively small. The nominal 95% confidence coverage rates are all close to 95%, except the 91.36% for the Exponential form and 91.75% for the Gompertz form for sample size 300.

3.5 Application to the BRAiNS Study

Subjects included in the current study (n=531) were assessed at least two times and comprise those included in a previous report [24]. All subjects were cognitively normal at study entry. The mean baseline age of these participants was 72.6 ± 7.5 years. The mean number of cognitive assessments for the cohort was 10.3 ± 4.1. During this follow-up period participants made transitions into MCI and or dementia, while many others died before such transitions.

The frequency of each type of transition is provided in Table 3.2. Note that over one-third of the subjects (35.6%) are still at risk for a serious cognitive impairment, while another third (35.8%) died before converting to a clinical MCI or dementia state. Another 19 subjects with MCI (3.6%) died before converting to dementia. Also, 105 (19.8%) of the participants transitioned to clinical MCI during follow-up, and 31 of these remain at risk for a dementia or death. Finally, 88 participants (16.6%) developed dementia during

follow-up, with 52 of these transitioning directly from normal to dementia between successive cognitive assessments.

Risk factors of interest here include APOE4 (Apolipoprotein E-4 allele) status, female gender, and low education (coded as 12 years or less, or more than 12 years). The subjects have an average of 16.0 ± 2.4 years of education. About 63.1% of the patients are female and about 30.4% have at least one APOE E-4 allele.

We applied the proposed Weibull model, as well as a time-homogenous model, and a piecewise-constant model to the BRAiNS data for comparisons. Time-homogenous model and piecewise-constant model have been discussed in detail by Jackson [29]. Since all participants were older than 60 years old at baseline, the time scale used in the Weibull model is the participant's age minus 60. In the piecewise-constant model, we divided the age into three periods, below 75, between 75 and 90, and above 90.

Figure 3.3 presents the baseline transition intensities estimated from three models. Solid line represents the intensity curves estimated by the proposed Weibull model. Dotted horizontal lines are estimated from the time-homogenous model, and the dash stepwise lines are from the piecewise-constant model. Here we could see both piecewise-constant model and Weibull model show that the transition intensities increase as participants get older.

Table 3.3 lists the hazard ratios and the corresponding 95% confidence interval (95% C.I.) for each covariate on each of the 5 transition paths by the three models mentioned above. The results of the hazard ratio estimates are close among these three models. Having at least one APOE4 (versus no APOE4) significantly increases the hazard rate for the transition from Normal to MCI, and cognitively normal females have

lower hazard of death than males. The proposed Weibull model also shows that having APOE4 would also increase the hazard ratio of transition from normal cognitive directly to dementia, which is consistent with the previous studies [2, 20, 22]. While, the time-homogenous model and piecewise-constant model failed to indicate the effect of APOE4 on transition from normal cognitive directly to dementia. The Akaike information criterion (AIC) statistics also show that the proposed Weibull model has the best fit among the three models, while the time-homogenous model has the worst fit. This further verifies that it is unrealistic to assume the time homogenous transition intensities in practice.

3.6 Discussion

Continuous-time multi-state models are useful in analyzing longitudinal event data. The regression models are simple and intuitive. All the characteristics of disease progression process could be modeled through the regression of intensity functions. The coefficients of the covariates have a similar log hazard ratio explanation as in survival models. Interval-censored data can be easily incorporated in the model, and it allows equally spaced longitudinal data in which the time interval between two consecutive longitudinal records varies.

We have presented non-homogeneous Markov models with incorporation of Weibull distribution to analyze multi-state longitudinal data. Our Weibull assumption allows us to have non-homogenous hazards, which is more appropriate for most applications than the widely used homogenous model. Unlike many other non-homogenous models, we are still able to construct the exact likelihood function through

our unique model structure, combining Weibull type non-homogenous and homogenous hazards.

Another advantage of our continuous Weibull model is that it allows us to fit both right censored and interval-censored data easily. In practice the mixed discrete and continuous pattern of observational data is very common in chronic disease studies. Patients are scheduled to visit the hospital at some pre-specified time points, so the exact transition times are interval-censored. In most cases, the death time is known exactly, but the state just before the death is unknown. Our continuous Weibull model works well for both situations.

One limitation of the proposed model is that the likelihood of our model does not have a closed form, which is needed to calculate the integrals. Multiple integrations were involved in the likelihood construction. The use of numerical integration solves our problem, but it reduces the estimation speed of our program, since hundreds of iterations are needed for each integration calculation. A faster and more reliable numerical integration method might help.

One possible avenue for future work is verification of the model assumptions, such as the Markov assumption for transition intensities and the proportional hazards assumption on covariate effects. A Semi-Markov Model might be a possibility if the Markov assumption is violated.

In conclusion, exponential regression Markov models with incorporation of the Weibull distribution were developed to a four state model to model the effects of covariates on the natural history of chronic disease with dispensing constant hazard assumption and with necessity for data on interval and right censored cases. In addition to

Alzheimer disease, our non-homogenous Markov model with Weibull assumption can be easily applied to data for other chronic disease with or without interval cases.

Table 3.1 Simulation results of covariate effects for sample sizes 300 and 500

N	True Baseline	Coefficient (True value)	Bias	MSE	95% CR
300	Exponential	$\beta_{12}(2.0)$	-0.007	0.075	91.4%
		$\beta_{13}(1.5)$	0.020	0.082	93.9%
		$\beta_{14}(0.0)$	-0.006	0.144	95.5%
	Weibull	$\beta_{12}(2.0)$	0.034	0.076	94.0%
		$\beta_{13}(1.5)$	0.030	0.070	95.0%
		$\beta_{14}(0.0)$	-0.020	0.066	95.6%
	Gompertz	$\beta_{12}(2.0)$	-0.089	0.104	91.8%
		$\beta_{13}(1.5)$	-0.039	0.099	93.0%
		$\beta_{14}(0.0)$	-0.055	0.060	95.0%
500	Exponential	$\beta_{12}(2.0)$	-0.005	0.043	94.4%
		$\beta_{13}(1.5)$	0.007	0.046	94.0%
		$\beta_{14}(0.0)$	-0.001	0.085	96.7%
	Weibull	$\beta_{12}(2.0)$	0.015	0.038	96.5%
		$\beta_{13}(1.5)$	-0.004	0.044	95.5%
		$\beta_{14}(0.0)$	0.006	0.036	96.3%
	Gompertz	$\beta_{12}(2.0)$	-0.083	0.062	93.3%
		$\beta_{13}(1.5)$	-0.065	0.061	95.0%
		$\beta_{14}(0.0)$	-0.055	0.032	95.8%

Note: N--number of subjects, MSE--mean square error, 95% CP--95% confidence coverage rate.

Table 3.2 Observed transition frequency of each transition type

Transition Type	Frequency	Percent %
Normal → Normal	184	35.65
Normal → MCI	52	9.79
Normal → Dementia	19	3.58
Normal → Death	190	35.78
Normal → MCI → MCI	31	5.84
Normal → MCI → Dementia	36	6.78
Normal → MCI → Death	19	3.58
Total	531	100

Table 3.3 Hazard Ratio estimates of each covariate by three models

Risk Factor	Path	Hazard Ratio (95% C.I.)		
		Time homogeneous	Piece-wise constant	Weibull
APOE4	1->2	**1.73** (1.14, 2.62)	**1.90** (1.25, 2.87)	**1.90** (1.26, 2.85)
	1->3	1.75 (0.80, 3.85)	1.98 (0.90, 4.37)	**2.03** (1.08, 3.81)
	1->4	0.71 (0.48, 1.04)	0.78 (0.53, 1.15)	0.83 (0.57, 1.22)
	2->3	1.01 (0.48, 2.13)	0.98 (0.46, 2.08)	0.98 (0.48, 2.00)
	2->4	1.90 (0.76, 4.75)	1.86 (0.73, 4.72)	1.78 (0.70, 4.52)
Low Education	1->2	1.55 (0.90, 2.67)	1.54 (0.89, 2.67)	1.55 (0.91, 2.64)
	1->3	0.39 (0.04, 3.85)	0.31 (0.02, 5.08)	0.44 (0.10, 1.98)
	1->4	1.19 (0.74, 1.91)	1.14 (0.71, 1.82)	1.07 (0.67, 1.72)
	2->3	0.92 (0.33, 2.60)	0.97 (0.34, 2.78)	0.94 (0.34, 2.56)
	2->4	0.61 (0.15, 2.53)	0.59 (0.14, 2.51)	0.63 (0.15, 2.55)
Female	1->2	0.82 (0.55, 1.23)	0.85 (0.57, 1.27)	0.79 (0.53, 1.18)
	1->3	1.94 (0.78, 4.79)	1.67 (0.69, 4.07)	1.73 (0.84, 3.57)
	1->4	**0.71** (0.52, 0.97)	**0.65** (0.48, 0.89)	**0.67** (0.50, 0.92)
	2->3	1.66 (0.83, 3.35)	1.75 (0.80, 3.83)	1.63 (0.81, 3.25)
	2->4	1.17 (0.48, 2.88)	1.21 (0.46, 3.19)	1.15 (0.46, 2.89)
AIC*		3521.48	3346.88	3298.25

Bold number: significant at 0.05 level. * Akaike information criterion.

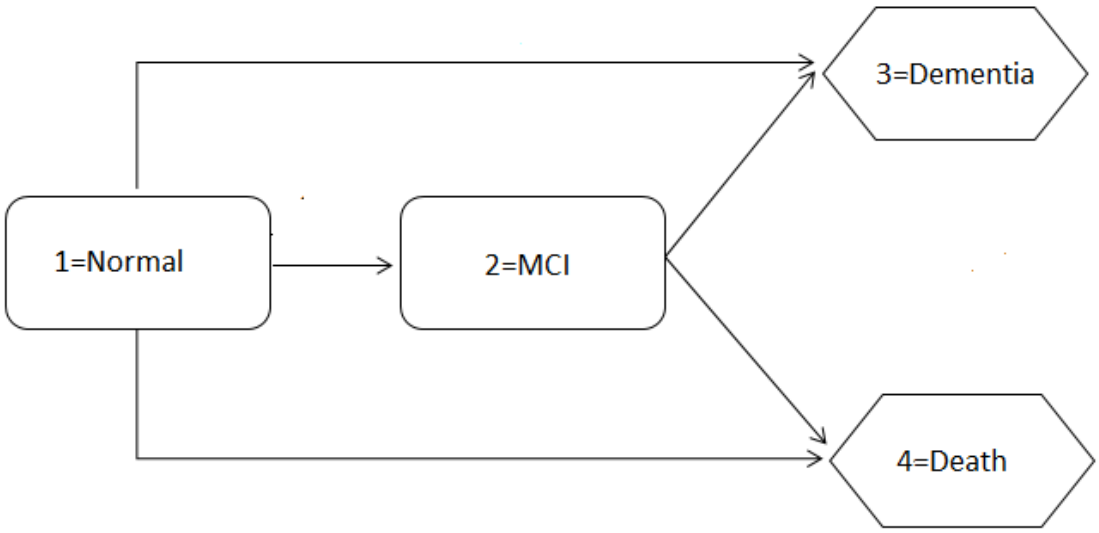

Figure 3.1 Transition flows of the four-state model

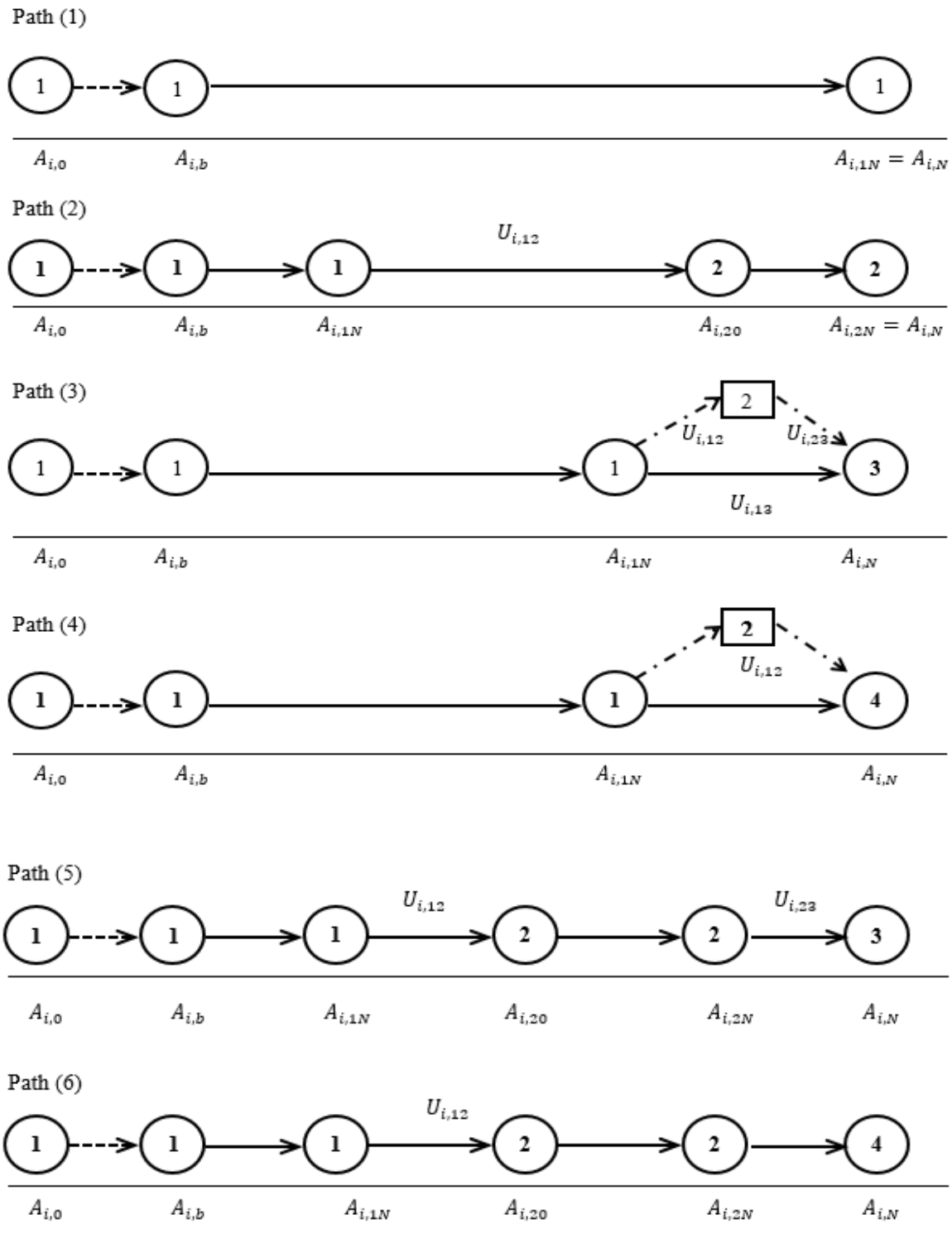

Figure 3.2 Possible observed transition paths

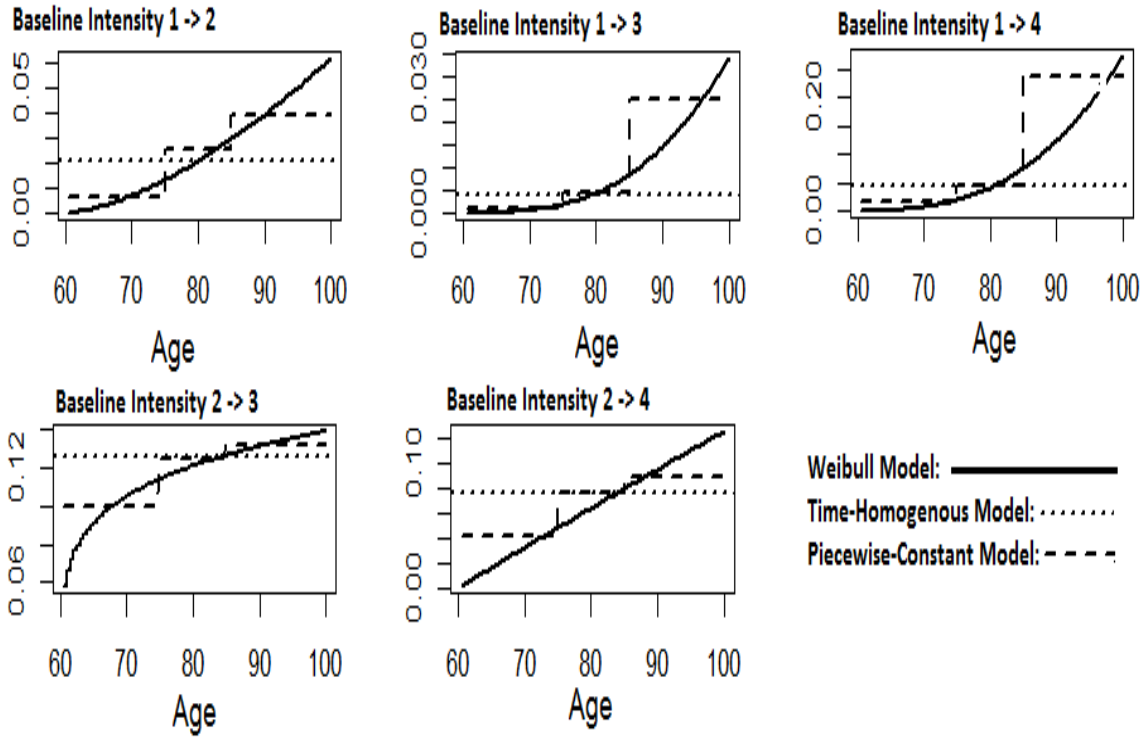

Figure 3.3 Baseline intensities estimated by three models for the BRAiNS data.

Chapter 4 A four-state Semi-Markov model with interval censored data and time-dependent covariates

4.1 Introduction

Longitudinal event-history data arises in many chronic disease studies, such as dementia [11], diabetes [39], HIV [40], cancer [41], liver cirrhosis [26], just to name a few. There are multiple possible states or stages in the process of the disease. For example, in the study of dementia, although dementia is the outcome of interest, study participants could first convert to clinically relevant states such as clinical mild cognitive impairment (MCI) for several years before finally entering a demented state. Participants in these longitudinal studies are usually observed over time. The outcome data consist of times of occurrence of transitions from one state to another state and the types of transitions that occur.

In the analysis of longitudinal event-history data, the first issue is to handle intermediate states and final absorbing states at the same time. The final absorbing states are usually the disease states of primary interest. However, the process of the disease will change dramatically if participants enter into a particular intermediate state. For example, in the development of dementia, participants with clinical MCI will have much higher risk converting to dementia than those without clinical MCI [20].

As an extension of survival models, Markov multi-state models [4-6, 42] enable researchers to investigate the transitions among multiple states at the same time. Two types of multi-state models are mainly used in practice. The multi-state Markov chain models are not appropriate when the longitudinal data are unequally spaced. Participants are usually assessed periodically, leading to interval censoring observations of transitions between the states. An alternative to handle interval

censored data is to model the transitions as a continuous-time Markov process. The process can be fully characterized by its transition intensities. In many applications, the Markov assumption might not be appropriate and may lead to biased conclusions. For instance, in the study of dementia, a person often first converts to a clinically relevant outcome such as clinical mild cognitive impairment (MCI) for several years before finally entering a demented state. The time spent at MCI seems to have a strong association to the future development of the dementia process. In this manuscript, we propose a continuous-time semi-Markov model to account for the effects of holding times on the future development of the disease process. The proposed model allows the transition intensities to be dependent on both the calendar time and the holding time the participant spends in the current state. This model can also handle left-truncation, interval-censored, right censored data. Both baseline and time-dependent covariates can be easily added into the model assuming the proportional hazards regression form. To facilitate the model building process, we also provide two model selection strategies and a graphic goodness-of-fit method based on prevalence counts [43].

The remaining of this chapter is structured as follows. In Section 4.2, we detail the model and the associated inference methods. In Section 4.3, we propose two model selection strategies to facilitate the model building process. In Section 4.4, a graphic goodness-of-fit method will be presented to check how the model fits the data. Results of the application to the BRAiNS data will be presented in Section 4.5. In the last section, we discuss the proposed method and outline some possible future directions.

4.2 The method

In this section, we introduce the notation and the likelihood function for the four-state semi-Markov model. The four-state model is an extension of the widely used "illness-death" model. We relax the commonly assumed Markov property, letting the future evolution of the process not only depend on the current state, but also on the entry time into the current state. Time dependent covariates can be easily incorporated into the model through proportional hazard properties for the transition intensities.

4.2.1 The Semi-Markov Framework

Different from a Markov process, the future of a semi-Markov process is not only dependent on the current state but also on the time the process entry into the current state. Let $X(t)$ be a continuous time semi-Markov process with a finite state space S $= \{1,2,3,4\}$. Define the transition probability from state h at time s to state j at time t given that the process entry into state h at time τ_h as

$$p_{hj}(s,t|\tau_h) = P(X(t) = j|X(s) = h, \tau_h) \;\; \tau_h < \mathrm{s} < \mathrm{t}.$$

The associated transition intensity has the following definition

$$\alpha_{hj}(t|\tau_h) = \begin{cases} \lim_{\Delta t \to 0} p_{hj}(t, t+\Delta t|\tau_h)/\Delta t\,, & j \neq h \\ -\sum_{k \neq h} \alpha_{hk}(t|\tau_h)\,, & j = h\,' \end{cases}$$

which represents the instantaneous hazard of transition from the current state h to state j at time t given the current state h and entry time τ_h when $h \neq j$.

The time scale t is the age of participants in this study. Since each subject is in the same initial state, normal, at baseline in our motivating example, we do not have a left truncation problem and simply assume a unique time (age 60) as the time origin for all subjects, which is the time we assumed participant entry into the state normal.

The strategy of dealing with initial time points is applied by Kryscio, et al. [22] and Kapetanakis, et al.[17]. Thus we have $\tau_1 = 60$.

4.2.2 Weibull Regression Model

Time dependent covariates $\boldsymbol{Z}(t)$ can be incorporated into the transition intensities using the proportional intensity regression model:

$$\alpha_{hj}(t|\tau_h) = \alpha_{hj,0}(t|\tau_h)\, exp\left(\boldsymbol{\beta}_{hj}^T \boldsymbol{Z}(t)\right), j \neq h\,.$$

Here $\alpha_{hj,0}(t|\tau_h)$ is called the baseline transition intensity. We assume the baseline intensities have the time reset property. Let $w = t - \tau_h$ be the holding time the participant has been in the current state, we have

$$\alpha_{hj,0}(t|\tau_h) = \begin{cases} \alpha_{hj,0}(t-\tau_h) = \alpha_{hj,0}(w), & t \geq \tau_h \\ 0 & t < \tau_h \end{cases}$$

We assume the baseline intensity functions have the Weibull form

$$\alpha_{hj,0}(w) = \lambda_{hj}\kappa_{hj} w^{\kappa_{hj}-1} = exp\big(\beta_{hj,0}\big)\kappa_{hj} w^{\kappa_{hj}-1}.$$

Here, the scale parameter is $\lambda_{hj} = exp\big(\beta_{hj,0}\big) > 0$ and the shape parameter is $\kappa_{hj} > 0$.

4.2.3 The Likelihood Function

Since we assume the target populations are all cognitively normal at the age of 60, and that participants who were not in the normal cognition state at their baseline age are left out of the dataset, the data are left truncated. And also we only observed the participants at discrete fixed time points, the entering times into state MCI and dementia are interval censored. The right censored data occurs when participants drop out of the study before they develop MCI or dementia, or is still in normal cognition or MCI at the end of the study. There are also possibilities that some transitions might not be observed. For example, if a participant was assessed as normal at this

assessment and then assessed as dementia at the next assessment, it is possible that he or she might have an observed transition from normal to MCI before finally converting to dementia sometime between assessments. We will consider all these cases when constructing the likelihood function.

To construct the likelihood, we let $A_{i,10}$ be the age before the beginning of the study at which participants first enter into the initial state (normal), here we have $A_{i,0} = \tau_1 = 60$. Also let $A_{i,b}$ be the age at baseline for participant i; $A_{i,1N}$ be the age at the last time participant i is observed in state 1 (normal); $A_{i,20}$ be the age at the first time the participant is observed in state 2 (MCI); $A_{i,2N}$ be the age at the first time the participant is observed in state 2; $A_{i,N}$ be the age at the last time participant i had an observation. Let $U_{i,hj}$ be the age when participant i transitions from state h to state j, for example, $U_{i,12}$ is the age at the time participant transition from state 1 to state 2 and $U_{i,23}$ is the age at the time participant transition from state 2 to state 3. These ages are all interval censored. However, the transition times to death, which are $U_{i,14}$ and $U_{i,24}$, are known exactly. In our case we have $U_{i,14} = A_{i,N}$ or $U_{i,24} = A_{i,N}$ if the last state recorded is death. Since every subject starts at state 1, and the process has no backward transitions, we have $\tau_2 = U_{i,12}$, which is interval censored: $A_{i,1N} < \tau_2 = U_{i,12} \leq A_{i,20}$. Thus we have

$$\alpha_{2j}(t|\tau_2) = \alpha_{2j}\big(t|U_{i,12}\big), t \geq U_{i,12}.$$

Before we construct the likelihood for the model, we denote two transition probabilities that will help us write the likelihood functions. First, we write $p_{11}\big(A_{i,b}, t\big) = P\big(X(t) = 1|X\big(A_{i,b}\big) = 1\big)$, which is the probability that the participant is still in state normal at time t given it was at state normal at baseline age $A_{i,b}$. We have

$$p_{11}(A_{i,b}, t) = exp\left(-\left(\Lambda_{i,1}(t) - \Lambda_{i,1}(A_{i,b})\right)\right).$$

Here, $\Lambda_{i,1}(t)$ is the cumulative hazard function of subject i for leaving state 1 and it has the form:

$$\Lambda_{i,1}(t) = \int_{A_{i,10}=60}^{t} \left(\alpha_{12}(u) + \alpha_{13}(u) + \alpha_{14}(u)\right)du$$

We also write $p_{22}(U_{i,12}, t) = P(X(t) = 2|X(U_{i,12}) = 2, \tau_2 = U_{i,12})$, which is the probability that the subject remains in state 2 given that it entered state 2 at time $U_{i,12}$. We have

$$p_{22}(U_{i,12}, t) = exp\left(-\int_{U_{i,12}}^{t} \left(\alpha_{23}(u|U_{h,12}) + \alpha_{24}(u|U_{h,12})\right) du\right)$$

There would be total 6 possible observed transition paths for a subject, as shown in Figure 3.2. We will discuss the likelihood construction below case by case.

Path (1): the patient has no transition during the study and is still in state 1 (normal) at the end of study. In this case, the likelihood for this patient would be

$$L_{i,1} = P(X(A_{i,N}) = 1|X(A_{i,b}) = 1) = p_{11}(A_{i,b}, A_{i,N})$$
$$= exp\left(-\left(\Lambda_{i,1}(A_{i,N}) - \Lambda_{i,1}(A_{i,b})\right)\right).$$

Path (2): the patient has one observed transition from state normal to state MCI and stays in state MCI at the end of study. In this case, we have

$$L_{i,2} = P(X(A_{i,1N}) = 1, X(A_{i,20}) = 2, X(A_{i,N}) = 2\ |X(A_{i,b}) = 1)$$
$$= \int_{A_{i,1N}}^{A_{i,20}} p_{11}(A_{i,b}, U_{i,12})\alpha_{12}(U_{i,12})p_{22}(U_{i,12}, A_{i,N})\, dU_{i,12}$$

Path (3): the patient has one observed transition from normal to state dementia. In this case, because of the interval-censoring, there could be two possible true paths, we need to take into account all the information available in such type of incomplete data. Scenario 1, the subject might have one transition from state normal directly to state dementia at time $U_{i,13}$. Scenario 2, the subject might have two transitions, first transition from state normal to state MCI at time $U_{i,12}$ then transition from state MCI to state dementia at time $U_{i,23}$. Thus, the likelihood of this path has two parts:

$$L_{i,3} = P\big(X(A_{i,1N}) = 1, X(A_{i,N}) = 3 | X(A_{i,b}) = 1\big)$$
$$= \int_{A_{i,1N}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,13})\alpha_{13}(U_{i,13})\, dU_{i,13}$$
$$+ \int_{A_{i,1N}}^{A_{i,N}} \int_{A_{i,1N}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,12})\alpha_{12}(U_{i,12})p_{22}(U_{i,12}, U_{i,23})\alpha_{23}(U_{i,23}|U_{i,12})dU_{i,12}dU_{i,23}$$

Path (4): subject i has one observed transition from state normal to death without dementia. Similar in Path (3), there could be two possible scenarios. The subject might have just one transition from state 1 directly to state 4 at time $A_{i,N}$; or it might have two transitions, first from state normal to state MCI at time $U_{i,12}$ then transition from state MCI to death at time $A_{i,N}$. Note that in BRAiNS data the exact age of death is recorded but the state just before death is unknown except dementia. Thus the subject might be at either state normal or state MCI before death. The likelihood for this path can be calculated as follows:

$$L_{i,4} = P\big(X(A_{i,1N}) = 1, X(A_{i,N}) = 4, U_{i,14} = A_{i,N} | X(A_{i,b}) = 1\big)$$
$$= p_{11}(A_{i,b}, U_{i,14})\alpha_{14}(U_{i,14})$$
$$+ \int_{A_{i,1N}}^{A_{i,N}} p_{11}(A_{i,b}, U_{i,12})\alpha_{12}(U_{i,12})p_{22}(U_{i,12}, A_{i,N})\alpha_{24}(A_{i,N}|U_{i,12})dU_{i,12}$$

Path (5): subject i has two observed transitions, first transition from state normal to state MCI at time $U_{i,12}$ $\left(A_{i,1N} < U_{i,12} \leq A_{i,20}\right)$ and then from state MIC to state dementia at time $U_{i,23}$ $\left(A_{i,2N} < U_{i,23} \leq A_{i,N}\right)$. In this case we have

$$L_{i,5} = P\left(X\left(A_{i,1N}\right) = 1, X\left(A_{i,20}\right) = 2, X\left(A_{i,2N}\right) = 2, X\left(A_{i,N}\right) = 3 \,|X\left(A_{i,b}\right) = 1\right)$$
$$= \int_{A_{i,1N}}^{A_{i,20}} \left(\int_{A_{i,2N}}^{A_{i,N}} p_{11}\left(A_{i,b}, U_{i,12}\right)\alpha_{12}\left(U_{i,12}\right)p_{22}\left(U_{i,12}, U_{i,23}\right)\alpha_{23}\left(U_{i,23}|U_{i,12}\right)dU_{i,23} \right) dU_{i,12}$$

Path (6): subject i has two observed transitions, the first is from state normal to state MCI at time $U_{i,12}$ $\left(A_{i,1N} < U_{i,12} \leq A_{i,20}\right)$ and the second is from state MCI to state death at time $A_{i,N}$. In this case we have

$$L_{i,6} = P\left(X\left(A_{i,1N}\right) = 1, X\left(A_{i,20}\right) = 2, X\left(A_{i,N} -\right) = 2\ , X\left(A_{i,N}\right) = 4|X\left(A_{i,b}\right) = 1\right)$$
$$= \int_{A_{i,1N}}^{A_{i,20}} p_{11}\left(A_{i,b}, U_{i,12}\right)\alpha_{12}\left(U_{i,12}\right)p_{22}\left(U_{i,12}, A_{i,N}\right)\alpha_{24}\left(A_{i,N}|U_{i,12}\right) dU_{i,12}$$

4.2.4 Parameter Estimations

The calculation of the likelihood function involves multiple integrals, which do not have closed forms. We implement the quasi-Monte Carlo (QMC) [44] method to approximate the likelihood. QMC will provide considerably better accuracy, with the expected integration error of the order of N^{-1} (N being the number of Halton sequence points from the integration space), to approximate the integrations of the likelihood function.[45].

Parameters contained in the model are the scale parameters $\lambda_{hj} = exp\left(\beta_{hj,0}\right)$, the shape parameters κ_{hj}, and the regression coefficients $\boldsymbol{\beta_{hj}}$, where $h < j, h \in \{1,2\}, j \in \{2,3,4\}$. Estimation and inference on these parameters can be achieved by

maximizing the likelihood function discussed above. We used “optim function” in R version 3.2.2 with the quasi-Newton method to maximize the log-likelihood function and to compute the numerically differentiated Hessian matrix. All programing was done in R.

4.3 Model Selection Strategy

In multi-state models, there are multiple possible transition paths and each covariate may have a different effect on different transition intensities. For example, in our four-state model there are 5 different transition intensity functions implying each covariate has up to 5 different coefficient parameters one per transition intensity. Thus model selection in multi-state models is more complicated than in other models such as linear models, logistical models, survival models, etc. In this section, we propose two model selection strategies for multi-state models. Both strategies help us to select the covariates and the coefficients on the associated transition path intensity functions and to determine the initial values for fitting the final selected model.

Strategy 1 is revised forward-backward step-wise selection method. The algorithm has the follow four steps:

Step 1. Fit a model with no covariate. Denote this model as $\mathcal{M}_0$.

Step 2. Add a single covariate $Z_1(t)$ into the model $\mathcal{M}_0$. Since $Z_1(t)$ would affect each transition with different coefficient, five parameters $(\beta_{12,1}, \beta_{13,1}, \beta_{14,1}, \beta_{23,1}, \beta_{24,1})$ are added into the model at the same time. Fit the model with the initial values for the baseline intensity parameters computed in Step 1. Next, apply a backward deletion. Parameter with the largest p value among the newly added parameters $(\beta_{12,1}, \beta_{13,1}, \beta_{14,1}, \beta_{23,1}, \beta_{24,1})$ is removed from the model. Refit the model and

repeat the above backward deletion algorithm until all covariate coefficients associated with $Z_1(t)$ are significant at level $p \leq 0.1$. Denote this model as $\mathcal{M}_1$.

Step 3. Add a second covariate $Z_2(t)$ into the previous model $\mathcal{M}_1$. Repeat Step 2. When applying the backward deletion algorithm in this step, we only delete the newly added parameters that are not significant at level $p \leq 0.1$. Parameters that are already in the previous model $\mathcal{M}_1$ are not removed even though they might be not significant at level $p \leq 0.1$. The resulting model after adding covariate $Z_2(t)$ and applying the backward deletion procedure is denoted as $\mathcal{M}_2$. Repeat the same procedure until all covariates are added in the model. Denote the resulting model as $\mathcal{M}_p$.

Step 4. Beginning from model $\mathcal{M}_p$, we apply a step-wise backward selection method. At each step, coefficient with the largest p value is removed from the model until all the coefficients are significant at level $p \leq 0.05$.

Strategy 2 is a two stage modeling technique. The first stage is a univariate modeling. The second stage is a multivariate modeling.

1. Univariate modeling –We calculated one model for each covariate. Models were said to be univariate, because only one factor was taken into account, even if it could influence a few transitions. At this stage, covariate coefficients are not significant at $p \leq 0.1$ level are removed from the univariate model in a step-fashion.

2. Multivariate modeling – All the previously selected significant covariates are included in the model. The vector of covariates were transition-specific. By a

backward step-wise deletion procedure, each coefficient with a p-value>0.05 is removed from the model.

4.4 Goodness of Fit

In this section, we provide a goodness-of-fit assessment for the proposed model by the means of prevalence counts [43]. Prevalence counts provide an informal empirical measure of state occupancy. If the model fits the data well, the expected state occupancies by the fitted model should be close to the observed state occupancies. By comparing the observed and expected prevalence counts, we would have a general goodness-of-fit assessment of the fitted model.

Let $\tilde{X}_i(t)$ be the observed process for the multi-state model. Since the process is observed only at some discrete time points, the transition time is interval censored for transition from state normal to state MCI and dementia and for transition from MCI to dementia. Thus, observed process $\tilde{X}_i(t)$ is unknown in the time interval with observed transitions. For example, subject i has an observed transition from state normal at time $A_{i,1N}$ to state dementia at time $A_{i,N}$, the value of $\tilde{X}_i(t)$ is unknown for time $t \in (A_{i,1N}, A_{i,N})$. Considering the proposed model is progressive, it is possible to interpolate the observed prevalence at any given time t. Titman, et al. [43] suggest assuming that the patient remains in the state they were in at the last observation. In this manuscript, we use the midpoint rule. As in the above example we have

$$\tilde{X}_i(t) = \begin{cases} 1 & t < (A_{i,1N} + A_{i,N})/2 \\ 3 & t \geq (A_{i,1N} + A_{i,N})/2 \end{cases}$$

Here, we assume that there are no unobserved transitions from state normal to state MCI in between the transition from state normal to state dementia.

Let $O_{1j}(t)$ be the observed prevalence counts in state j at time t among subjects started in state normal at baseline age. And let $O_{2j}(t)$ be the observed prevalence counts in state j at time t among subjects having a transition to state MCI. We have

$$O_{1j}(t) = \sum_{i=1}^{N} I\left(\tilde{X}_i(t) = j | \tilde{X}_i\left(A_{i,b}\right) = 1\right)\delta_{1i}(t)$$
$$O_{2j}(t) = \sum_{i=1}^{N} I\left(\tilde{X}_i(t) = j | \tilde{X}_i\left(U_{i,12}\right) = 2\right)\delta_{2i}(t)$$

Here, $\delta_{1i}(t)$ and $\delta_{2i}(t)$ are indicators of whether patient i was under observation at time t:

$$\delta_{1i}(t) = \begin{cases} 0, & t > A_{i,N} \; and \; \tilde{X}_i\left(A_{i,N}\right) \in (1,2) \\ 1, & otherwise \end{cases}$$

and

$$\delta_{2i}(t) = \begin{cases} 0, & t > A_{i,N} \; and \; \tilde{X}_i\left(A_{i,N}\right) \in (1,2) \\ 1, & A_{i,N} > t \geq U_{i,12} \end{cases}.$$

The calculation of $O_{2j}(t)$ is dependent on the transition time to state MCI, we assume $U_{i,12} = \left(A_{i,1N} + A_{i,20}\right)/2$ for the purpose of calculating the prevalence counts among these subjects. In the above observed prevalence counts, we assume that there are no unobserved transitions. The observed prevalence counts for state MCI would be under estimated, since we do not account for the possible unobserved transitions from state normal to state MCI for subjects with observed transition from normal directly to dementia or death.

The calculations of expected prevalence counts are straightforward. Denote $E_{1j}(t)$ and $E_{2j}(t)$ the expected prevalence counts in state j at time t among all subjects and among subjects having a transition into MCI respectively. We have

$$E_{1j}(t) = \sum_{i=1}^{N} \hat{P}\big(X_i(t) = j | X_i(A_{i,b}) = 1\big)\delta_{1i}(t)$$

$$E_{2j}(t) = \sum_{i=1}^{N} \hat{P}\big(X_i(t) = j | X_i(U_{i,12}) = 2\big)\delta_{2i}(t)$$

Here the expected transition probability from state normal at baseline to state j at time t, $\hat{P}\big(X_i(t) = j | X_i(A_{i,b}) = 1\big)$, and the expected transition probability from state MCI at time $U_{i,12}$ to state j at time t, $\hat{P}\big(X_i(t) = j | X_i(U_{i,12}) = 2\big)$ can be calculated by using in the estimated model parameters and covariate values. Here we assume $U_{i,12} = \big(A_{i,1N} + A_{i,20}\big)/2$.

A comparison of the observed prevalence counts and the expected prevalence counts by the fitted model can be made by plotting the observed prevalence and expected prevalence functions on the same graph.

4.5 Application

BRAiNS is a longitudinal cohort of 1,030 older participants at the University of Kentucky's Alzheimer's disease Center (UK ADC) [36]. Participants consent to extensive annual cognitive and clinical examinations. The sample included in this study consists of 531 participants, who were assessed at least two times, and were at least 60 years old at baseline. All subjects were cognitively intact at study entry. The baseline age for the sample is 73.2 (STD=7.4) years. We have 6 possible transition paths among the four states under consideration. Table 4.1 presents the frequency and percentage for each observed transition path.

The list of factors to be examined as potential risks for transitions among the states were selected by matching factors reported in the literature with the data

elements collected on participants in the BRAiNS cohort. The factors examined and entered as indicator variables in the statistical models below are: APOE4 carrier status (with and without an ε4 allele), female gender, low education (defined as high school or less), family history of dementia among 1st degree relatives, baseline current smoker and presence of Type II diabetes. These covariates are all baseline covariates. See Table 4.2 for the summary of these covariates. A time dependent covariate Age Group will also be added to the transition intensity functions from state MCI to dementia and death. Age Group has two levels and it is defined as follows:

$$Age\ Group(t) = \begin{cases} 0, & t < 82.5 \\ 1, & t \geq 82.5 \end{cases}.$$

Here, the time scale t is the age and the cut point 82.5 is set to be around the median value of midpoint of the time interval who transition from state normal to state MCI.

We applied both model selection strategies presented in Section 4, which resulted in the same final model. The parameter estimates of the final model are listed in Table 4.3. The time dependent covariate Age Group is not significant on the intensity from clinical MCI to dementia, but it is significant on the transition path from MCI to death. As expected, older subjects have higher mortality rate for these have been in clinical MCI. Having at least one APOE4 allele increases the log hazard of transition from NSI into clinical MCI. Female gender has lower hazard for the transition from NSI to death than male. And baseline smoker increase the intensity rate for both the transition from NSI to death and from clinical MCI to dementia. Family history of dementia and baseline type II diabetes are not significant in this model.

In left panel of Figure 4.3, we plotted the baseline transition intensities against age for the three transitions from NSI to clinical MCI, from NSI to dementia and from NSI to death. The dark solid line represents the intensity function for the transition from NSI to clinical MCI. The dark dot line represents the intensity for the transition path from NSI to dementia. And the light dot line represents the intensity of the transition from NSI to death. The plots show that the intensities for the transition from NSI to clinical MCI and death increase steadily with the later has a higher increase rate; the transition intensity for the path from NSI directly to dementia remain flat under age 80, then it begins to increase as subject gets older.

In the right panel of Figure 4.3, we plotted the baseline transition intensities from clinical MCI to either dementia or death against the time since subject first went into clinical MCI state. The dark solid line represents the transition intensity of path from clinical MCI to dementia. The dark dot line represents the transition intensity of the path from clinical MCI to death for these older than 82.5 years of age. And the light dot line represent the intensity of the path form clinical MCI to death for these who are younger than 82.5 years of age. In these plots, the intensity rate of the transition from clinical MCI to dementia is relatively flat against the years the subject has spent in the MCI state. The intensities of both age groups for transition from MCI to death increase steadily as subjects spend more time on state MCI. As we noted in the plots, older subjects have higher increased rates than younger subjects.

We check the goodness-of-fit of the model by the prevalence plots discussed in Section 5. The goodness of fit plots are presented in Figure 4.4 and Figure 4.5. The dots are the observed prevalence counts, and the solid dark lines are the expected prevalence counts in these plots. Figure 4.4 presents the prevalence counts for all subjects started at state NSI. Figure 4.5 presents the prevalence counts for those

subjects having an observed transition to MCI. Except for the prevalence counts for MCI in Figure 4.4, there is a good agreement between the expected prevalence and the observed prevalence, which shows the proposed model is a good fit to the data. One possible reason for the disagreement between observed and expected prevalence of MCI for those subjects started at NSI state in Figure 4.4 is that there might be unobserved transitions from NSI to MCI due to the interval censoring.

4.6 Discussion

In this chapter we proposed a four-state continuous-time semi-Markov model applicable for left truncated, interval and right censored data. The proposed model also allows time-dependent covariates. Two model selection strategies are proposed to help select the "best" model that both fits the data and has a manageable number of parameters.

Despite this connection between our semi-Markov model and the one proposed by Foucher, et al. [16] and Kryscio, et al. [22], the modeling techniques are different. In their models, the semi-Markov process was modeled through two separate parts. The first part models one-step transition probabilities using standard logistic models; and the second part models the hold times given the transition paths. We need two coefficient parameters for each covariate on each possible transition path; one coefficient assesses its effect on transition probability and the other one assesses its effects on the holding time if that transition occurs. In our model, we modeled the process through the transition intensities. We allow the transition intensities to be dependent on both the calendar time and the holding time the subject has been in the state. We need only one parameter for each covariate on each transition path. And our model also allows time-dependent covariates.

The graphic goodness-of-fit method proposed in Section 4.5 can only be used a rough measurement of how will the model fits the data, since the true observed prevalence counts are unknown because of the interval-censoring. The observed prevalence counts calculated using the mid-point rule are usually under-estimated for intermediate states, since the unobserved transitions into these states are ignored. More sensitive goodness-of-fit tools are needed to for multi-state models.

Table 4.1 Frequency of each transition type

Observed Path	N	Percent
NSI -> NSI	184	34.65
NSI -> MCI	50	9.42
NSI -> Dementia	52	9.79
NSI -> Death	190	35.78
NSI -> MCI -> Dementia	36	6.78
NSI -> MCI -> Death	19	3.58
All	531	100

Table 4.2 Summary of the fixed covariates

Baseline Characteristic	N	Percent
APOE4	160	30.3
Low Education	187	35
Female	334	63.1
Family history of dementia	214	40.3
Baseline smoker	49	9.2
Type II diabetes	44	8.3

Table 4.3 Parameter estimates for the four-state semi-Markov model

Covariate	Regression coefficient	Estimation	S.E.	P value
Shape parameters	κ_{12}	2.467	0.256	0.000
	κ_{13}	5.613	1.472	0.000
	κ_{14}	3.869	0.261	0.000
	κ_{23}	1.119	0.121	0.000
	κ_{24}	2.606	0.455	0.000
Model intercepts	$\beta_{12,0}$	-9.138	0.892	0.000
	$\beta_{13,0}$	-21.554	5.283	0.000
	$\beta_{14,0}$	-13.274	0.907	0.000
	$\beta_{23,0}$	-1.732	0.233	0.000
	$\beta_{24,0}$	-6.089	1.025	0.000
Age group	$\beta_{24,G}$	1.349	0.659	0.041
APOE4	$\beta_{12,A}$	0.717	0.182	0.000
Female	$\beta_{14,F}$	-0.328	0.147	0.026
Baseline Smoker	$\beta_{14,S}$	0.878	0.210	0.000
	$\beta_{23,S}$	1.448	0.611	0.018

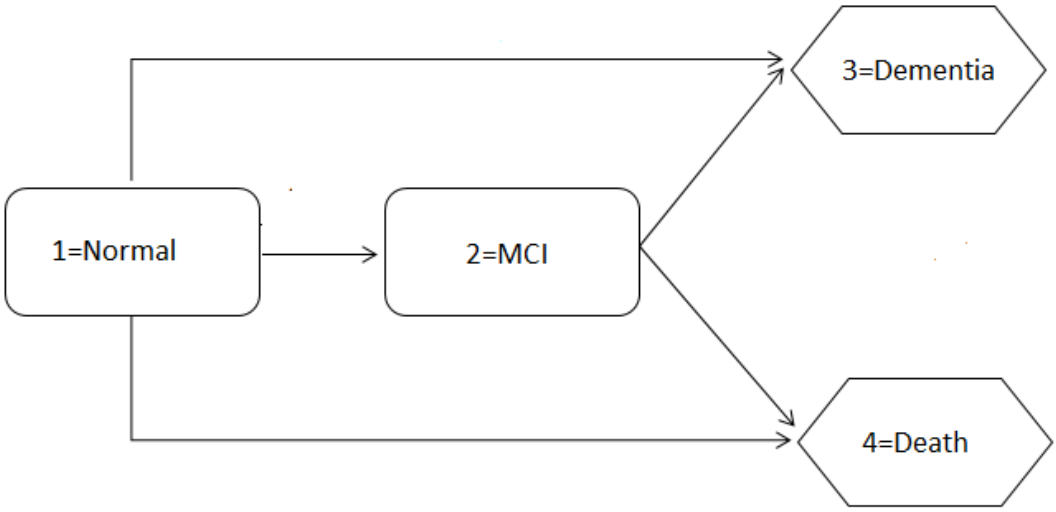

Figure 4.1 Model structure of the four-state model

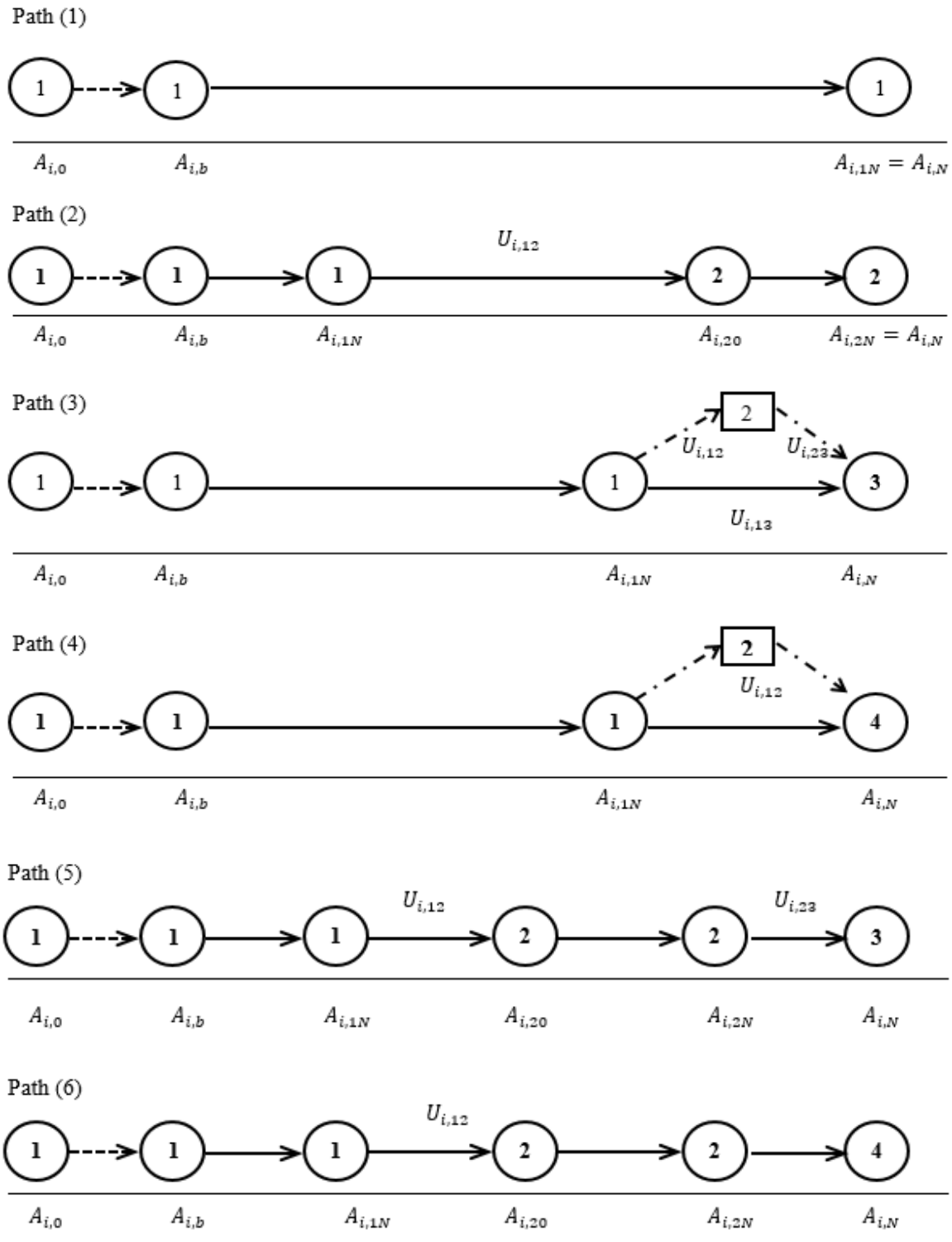

Figure 4.2 Possible observed transition path of a participant

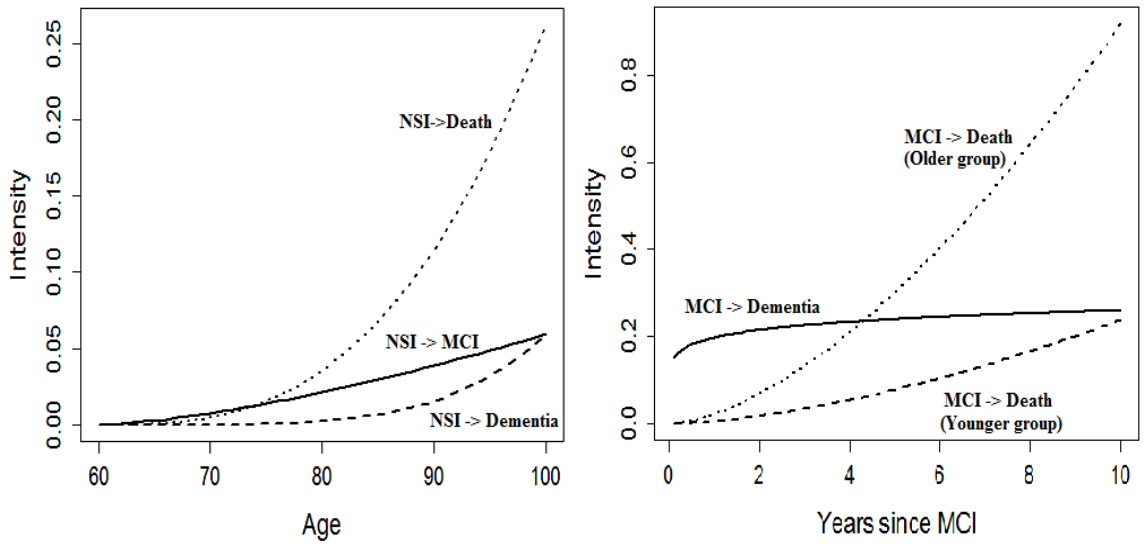

Figure 4.3 Baseline transition intensity plots.

Left panel: dark solid line-from NSI to clinical MCI; dark dot line-from NSI to dementia; light dot-from NSI to death.

Right panel: dark solid line-from clinical MCI to dementia; dark dot line-from clinical MCI to death (older group); light dot line-form clinical MCI to death (younger group)

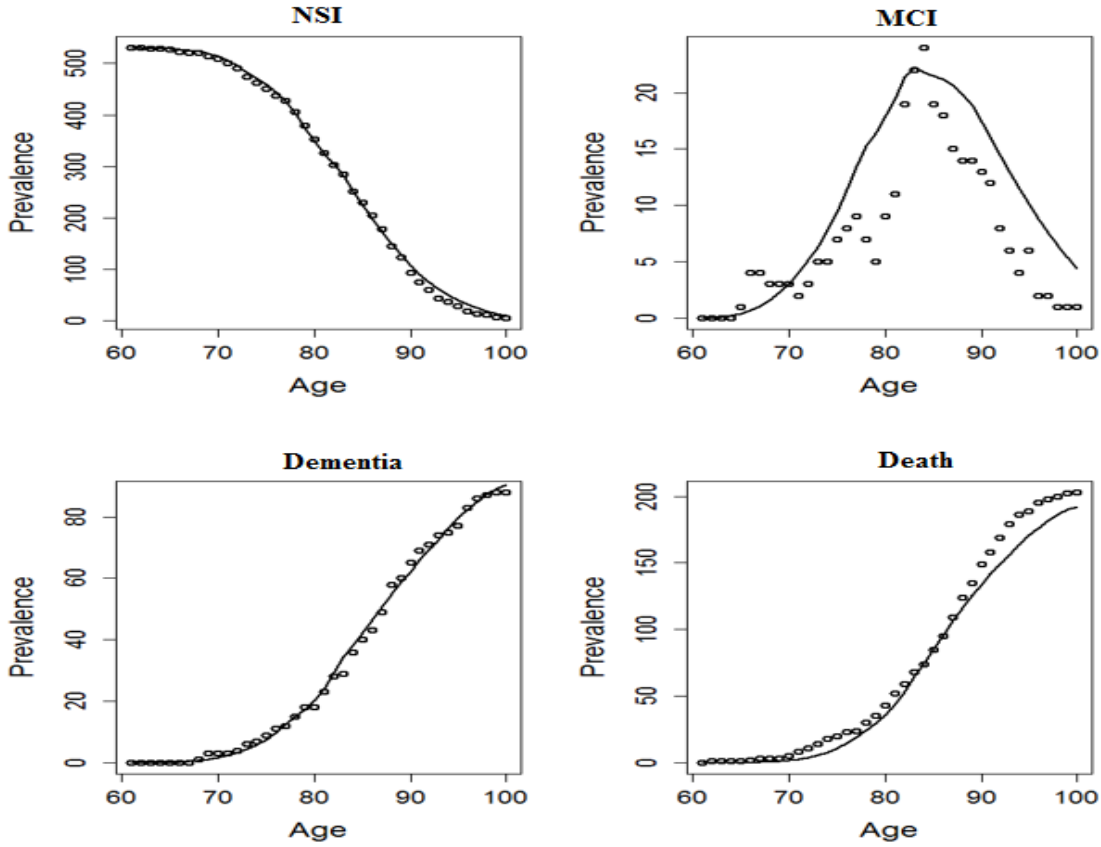

Figure 4.4: Prevalence plots for all the subjects started at NSI at 60 years old. Dots: Observed prevalence counts; Lines: expected prevalence counts.

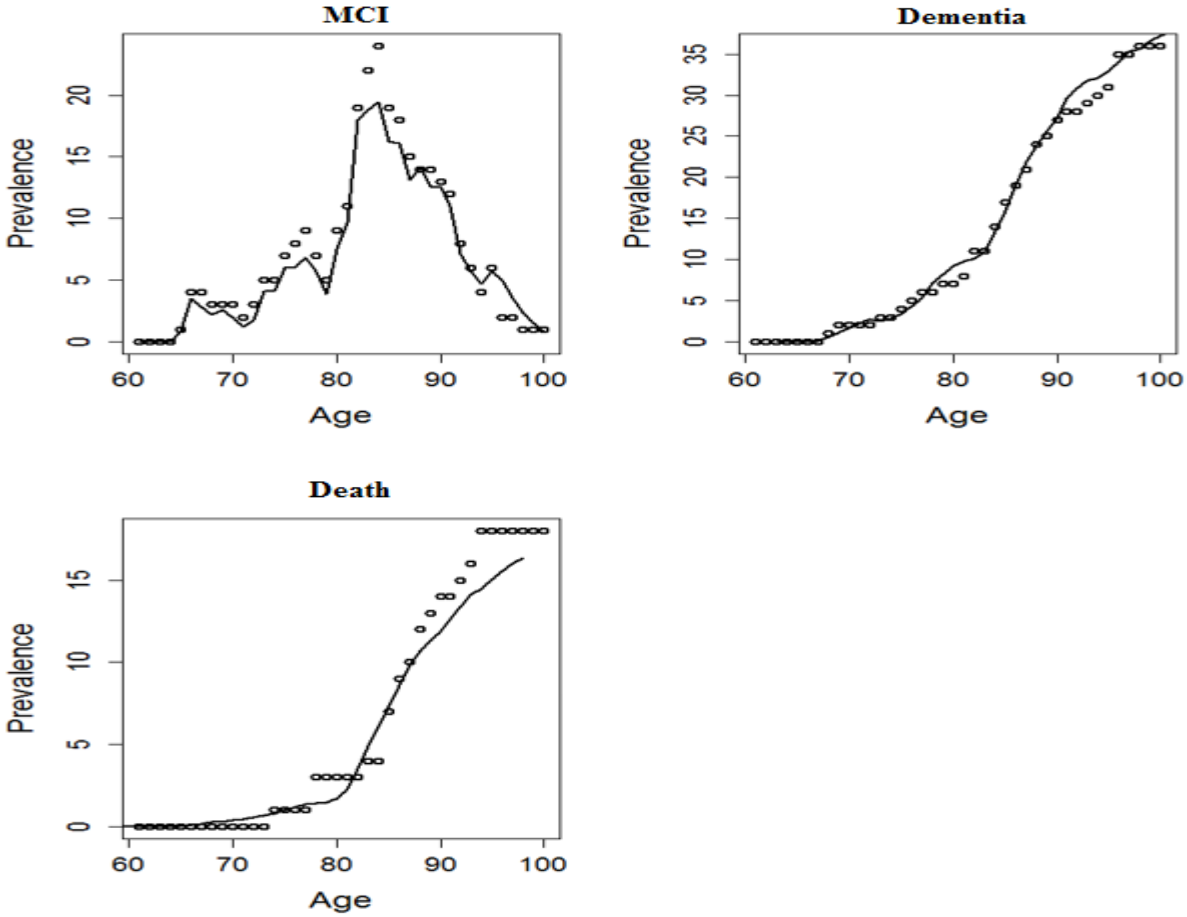

Figure 4.5 Prevalence plots for these subjects having an observed transition to MCI

Chapter 5 Discussions and Future Research

In dementia studies, clinical assessments of a participants' cognitive status is taken periodically at discrete visit time points; and the cognitive status can be classified into different states. This type of longitudinal data is commonly modeled by discrete-time Markov chain models. Since the clinical assessments are taken periodically at discrete time points, the transition times from one state to another state are interval-censored. Another problem caused by the discrete-time observation scheme is that some transitions might not be observed. In this dissertation, we explored the use of continuous-time multi-state models to analyze this type of longitudinal data raised in many chronic disease studies.

First, we compared the two types of multi-state models, discrete-time Markov chain model and continuous-time Markov process model. Our study showed that when the data are equally-spaced the two types models perform equally well. However, when the data are not equally-spaced, the continuous-time Markov process model has better performance than the discrete-time Markov chain model. The Markov chain model is biased when the data is unequally-spaced. Thus, our recommendation is that when the data are equally-spaced, either type of multi-state model can be used. The discrete-time Markov chain model might be more attractive to some researchers since it can be solved through standard statistical software. When the data are unequally-spaced, the continuous-time process model is recommended.

Calculations involved in the general continuous-time Markov process model could be very complex. Time-homogenous assumption is often used to simplify these calculations. However, the time-homogenous assumption is not appropriate in some cases. As we known, the hazards of developing dementia and death are heavily depended on a participants' age. Older people generally have higher risk of

developing dementia or risk of mortality. In this dissertation, we propose a Weibull Markov four-state model with two transient states and two absorbing states. By using the Weibull hazard form, we were able to model a variety of shapes for the transition intensities. This methodology can be easily generalized to models with more states as long as there are no backward transitions. With backward transitions, the likelihood calculation under the Weibull model will be complex, thus more powerful numerical calculation method should be developed in this case. This would be an interesting possible future research topic.

The Markov assumption is very common in the literature. Relaxing the Markov assumption is a challenging but also an important topic. As we know, in some chronic diseases studies the past history of the disease would have large impact on the future development of the disease. For example, beside the participant's age, the time that participant has stayed in state MCI would also have an impact on its transition hazards to dementia and death. In this dissertation, we proposed a semi-Markov model to allow the disease process not only depended on the participants' age but also on the time they have stayed on the current state. The importance of this model is that it helps us understand the participants' transition hazards and future transition probabilities based on their time stayed in the current state, thus we can treat participants differently according to the time they have stayed in some critical states, i.e. MCI.

Both the Markov model and semi-Markov model we presented in this dissertation requires the data contains no backward transitions. Treating backward transitions in a discrete-time observation scheme in Markov model and semi-Markov model is challenging in two ways. First, transition times are interval-censored. With interval-censored data, multiple integrations are involved in likelihood calculations.

Second, there are unobserved transitions. In a multi-state model with backward transitions, theoretically there are infinite numbers of unobserved transitions.

Appendices

A. SAS codes for Chapter 2

```
/*********************************************************************/
/*1 Simulation Study*/
/*********************************************************************/
PROC IML; reset storage= P1Sim.W1Sim;
START TranPD(parms,z);
   P=J(4,4,0);
   int1=parms[1];int2=parms[2];int3=parms[3];
   ps1=parms[4];ps2=parms[5];ps3=parms[6];
   z1=parms[7];z2=parms[8];z3=parms[9]; ps=1;
   pexp1=exp(int1+ps1*ps+z1*z); pexp2=exp(int2+ps2*ps+z2*z);
   pexp3=exp(int3+ps3*ps+z3*z);
   P[1,2]=pexp1/(1+pexp1+pexp2+pexp3); P[1,3]=pexp2/(1+pexp1+pexp2+pexp3);
   P[1,4]=pexp3/(1+pexp1+pexp2+pexp3); P[1,1]=1/(1+pexp1+pexp2+pexp3);
   ps=0;
   pexp1=exp(int1+ps1*ps+z1*z); pexp2=exp(int2+ps2*ps+z2*z);
   pexp3=exp(int3+ps3*ps+z3*z);
   P[2,2]=pexp1/(1+pexp1+pexp2+pexp3); P[2,3]=pexp2/(1+pexp1+pexp2+pexp3);
   P[2,4]=pexp3/(1+pexp1+pexp2+pexp3); P[2,1]=1/(1+pexp1+pexp2+pexp3);
   return(P);
FINISH TranPD;

START TranPC(parms,z);
   Q=J(4,4,0);
   Q[1,2]=exp(parms[1]+parms[7]*z); Q[1,3]=exp(parms[2]+parms[8]*z);
   Q[1,4]=exp(parms[3]+parms[9]*z); Q[1,1]=-Q[1,2]-Q[1,3]-Q[1,4];
   Q[2,1]=exp(parms[4]+parms[10]*z); Q[2,3]=exp(parms[5]+parms[11]*z);
   Q[2,4]=exp(parms[6]+parms[12]*z); Q[2,2]=-Q[2,1]-Q[2,3]-Q[2,4];
   A=teigvec(Q); V=teigval(Q);  D=diag(exp(V[,1])); P=A*D*inv(A); return(P);
FINISH TranPC;

START TrueP(parms,z);
   Q=J(4,4,0); Q[1,2]=exp(log(parms[1])+parms[7]*z);
   Q[1,3]=exp(log(parms[2])+parms[8]*z); Q[1,4]=exp(log(parms[3])+parms[9]*z);
   Q[1,1]=-Q[1,2]-Q[1,3]-Q[1,4];
   Q[2,1]=exp(log(parms[4])+parms[10]*z);  Q[2,3]=exp(log(parms[5])+parms[11]*z);
   Q[2,4]=exp(log(parms[6])+parms[12]*z); Q[2,2]=-Q[2,1]-Q[2,3]-Q[2,4];
   A=teigvec(Q); V=teigval(Q);  D=diag(exp(V[,1])); P=A*D*inv(A); return(P);
FINISH TrueP;

START W1Sim(type, Nsim);
   row=J(1,12,0); rows=J(1,12,0);
   do iSim=1 to Nsim;
     Submit type;
       Data W1Sim1;
          do subject=1 to 500;
            z=RAND('BERNOULLI',0.4); time=0; output;
```

```
            do i=1 to 15;
              if &type=1 then duration=1; else if &type=2 then
              duration=RAND('NORMAL',1,0.5);
               else if &type=3 then duration=RAND('NORMAL',1,1.5);
               if duration<0.25 then duration=0.25; time=time+duration; output;
             end;
           end;
         Run;
       Endsubmit;
       Run ExportDataSetToR("W1Sim1", "W1Sim1" );
       Submit /R;
       library("msm")
       qmatrix <- rbind(c(0, 0.25,  0.03, 0.05), c(0.2,0, 0.15, 0.05 ), c(0,0,0,0),
          c(0,0,0,0))
       W1Sim2=simmulti.msm(W1Sim1, qmatrix,
         covariates=list(z = c(0.5,-0.2,0,      -0.3, 0.15, 0)), death=4)
        ObsL.msm <-msm(state ~ time, subject=subject, data =W1Sim2,
          covariates = list("1-2" =~z, "1-3"=~z,  "2-1"=~z,"2-3"=~z),qmatrix =qmatrix,
          death=4,center=FALSE, method = "BFGS", control = list(fnscale = 4000,
          maxit = 10000))
        EstR=ObsL.msm$estimates
      Endsubmit;
      Run ImportDatasetFromR("W1Sim2", "W1Sim2" ); Run
     ImportMatrixFromR(EstC, "EstR" );
     Submit;
       Data W1Sim3; set W1Sim2; by subject time; Pstate=lag(state);
         if first.subject ne 1 then output; keep subject time Z state pstate; Run;
       Data W1Sim4; set W1Sim3;if state=1 then state=5;if pstate=2 then pstate=0; Run;
       ods select none;
       Proc CATMOD data=W1Sim4; direct pstate Z; model state=pstate Z;
       ods output Estimates=EstD; Run;
       ods select all;
     Endsubmit;
     Use EstD; read all var{estimate} into parms; close EstD;
     PD0=TranPD(parms,0);row[1]=type;row[2]=isim;row[3]=1;row[4]=0;row[5:8]=
     pd0[1,];row[9:12]= pd0[2,];
     if iSim=1 then rows=row; else rows=rows//row;
     PD1=TranPD(parms,1); row[1]=type; row[2]=iSim; row[3]=1;
     row[4]=1; row[5:8]=PD1[1,];row[9:12]=PD1[2,];rows=rows//row;
     PC0=TranPC(EstC,0); row[1]=type; row[2]=iSim; row[3]=2; row[4]=0;
     row[5:8]=PC0[1,]; row[9:12]=PC0[2,];rows=rows//row;
     PC1=TranPC(EstC,1);row[1]=type; row[2]=iSim; row[3]=2; row[4]=1;
     row[5:8]=PC1[1,]; row[9:12]=PC1[2,];rows=rows//row;
   End;
   varnames={"Type" "IDSim" "Method" "Z" "P11" "P12" "P13" "P14" "P21" "P22"
   "P23" "P24"};
   create W1Sim from rows[colname=varNames];append from rows; close W1Sim;
FINISH W1Sim;
STORE module=_all_;
Quit;
```

```
PROC IML;
   reset storage= P1Sim.W1Sim;  /* set location for storage */ load module=_all_;
   call W1Sim(1,1000); Submit; Data W1Sim1; set W1Sim; Run; Endsubmit;
   call W1Sim(2,1000); Submit; Data W1Sim2; set W1Sim; Run; Endsubmit;
   call W1Sim(3,1000); Submit; Data W1Sim3; set W1Sim; Run; Endsubmit;
Quit;


/*************************************************************************/
/*2 Application to the Nun's data*/
/*************************************************************************/
/*Discrete-time MSM*/
Data Nun_p1; set Nun_4state; where priorstate=1; run;
PROC logistic data=nun_p1;model currentstate(ref="1")=Bage apoe4/link=glogit;run;
Data Nun_p2; set Nun_4state; where priorstate=2; run;
PROC logistic data=nun_p2;model currentstate(ref="2")=Bage apoe4/link=glogit;run;
Data Nun_p3; set Nun_4state; where priorstate=3; run;
PROC logistic data=nun_p3;model currentstate(ref="3")=Bage apoe4/link=glogit;run;

/*Continous-time MSM*/
PROC IML;
START loglhm(parms) global(dataset, error);
n=nrow(dataset);Q=J(4,4,0); logLike=.0;
do i=1 to n;
   first_id=dataset[i,1];page=dataset[i,2];cage=dataset[i,3];pstate=dataset[i,4];
   cstate=dataset[i,5];  bage=dataset[i,6]; apoe4=dataset[i,7];
   Q[1,2]=exp(parms[1]+parms[8]*bage+parms[15]*apoe4);
   Q[1,3]=exp(parms[2]+parms[9]*bage+parms[16]*apoe4);
   Q[1,4]=exp(parms[3]+parms[10]*bage+parms[17]*apoe4);
   Q[2,1]=exp(parms[4]+parms[11]*bage+parms[18]*apoe4);
   Q[2,3]=exp(parms[5]+parms[12]*bage+parms[19]*apoe4);
   Q[2,4]=exp(parms[6]+parms[13]*bage+parms[20]*apoe4);
   Q[3,4]=exp(parms[7]+parms[14]*bage+parms[21]*apoe4);
   Q[1,1]=-Q[1,2]-Q[1,3]-Q[1,4]; Q[2,2]=-Q[2,1]-Q[2,3]-Q[2,4]; Q[3,3]=-Q[3,4];
   A=teigvec(Q); V=teigval(Q);D=diag(exp(V[,1]*(cage-page))); P=A*D*inv(A);
   if cstate=4 then Li=P[pstate,1]*Q[1,4]+P[pstate,2]*Q[2,4]+P[pstate,3]*Q[3,4];
   else Li=P[pstate,cstate]; if Li<=0 then error=error+1;
   else logLike=logLike+log(Li);
end; return(-logLike);
FINISH loglhm;

USE Nun;
Read all var{ID priorage  currentage  priorstate  currentstate bage apoe4} into dataset;
CLOSE Nun;
h0={-2.8451 -4.6193 -3.8569 -1.7184 -1.8959 -1.2406 -1.6418
0.0887   0.1269   0.0638   0.0100   0.0381   -0.0211   0.0364
0.4383   0.0941   0.3494   -0.6053   0.4893   -0.5624   0.0110};
error=0; call nlpnra(rc,xres,"loglhm",h0); estimate=xres` ; call
nlpfdd(f,g,hes1,"loglhm",estimate);
cov=inv(hes1); stderr=sqrt(abs(vecdiag(cov)));
```

```
z=abs(estimate/stderr);p=2*(1-probnorm(z));
print error f; print estimate stderr p; print cov;
QUIT;

/*****************************************************************************/
/*3 Transition probability plots*/
/*****************************************************************************/
%MACRO TransP(Age,Bage,APOE4);
PROC IML;
START TranPMD(parms, Age, Bage, APOE4);
   P=J(4,4,0);
   /*Bage APOE4 age p11 p12 p13 p14 p21 p22 p23 p24 p33 p34*/
   rows={0 0 0 0 0 0 0 0 0 0 0 0 0};
   pexp11=1;
   pexp12=exp(parms[1]+parms[8]*bage+parms[15]*apoe4);
   pexp13=exp(parms[2]+parms[9]*bage+parms[16]*apoe4);
   pexp14=exp(parms[3]+parms[10]*bage+parms[17]*apoe4);
   pexp22=1;
   pexp21=exp(parms[4]+parms[11]*bage+parms[18]*apoe4);
   pexp23=exp(parms[5]+parms[12]*bage+parms[19]*apoe4);
   pexp24=exp(parms[6]+parms[13]*bage+parms[20]*apoe4);
   pexp33=1;
   pexp34=exp(parms[7]+parms[14]*bage+parms[21]*apoe4);
   pexp1=pexp11+pexp12+pexp13+pexp14;
   P[1,1]=pexp11/pexp1; P[1,2]=pexp12/pexp1; P[1,3]=pexp13/pexp1;
   P[1,4]=pexp14/pexp1;
   pexp2=pexp21+pexp22+pexp23+pexp24;
   P[2,1]=pexp21/pexp2; P[2,2]=pexp22/pexp2; P[2,3]=pexp23/pexp2;
   P[2,4]=pexp24/pexp2;
   P[3,3]=pexp33/(pexp33+pexp34); P[3,4]=pexp34/(pexp33+pexp34); P[4,4]=1;
   do i=0 to Age;
      if i=0 then TPM=I(4); else TPM=TPM*P;
      row={0 0 0 0 0 0 0 0 0 0 0 0 0}; row[1]=Bage; row[2]=APOE4; row[3]=i;
      row[4]=TPM[1,1]; row[5]=TPM[1,2]; row[6]=TPM[1,3]; row[7]=TPM[1,4];
      row[8]=TPM[2,1]; row[9]=TPM[2,2]; row[10]=TPM[2,3]; row[11]=TPM[2,4];
      row[12]=TPM[3,3]; row[13]=TPM[3,4];rows=rows//row;
   end;
   varnames={"Bage" "APOE4" "Age" "P11D" "P12D" "P13D" "P14D" "P21D"
      "P22D" "P23D" "P24D" "P33D" "P34D"};
   create TranPMD from rows[COLNAME=varNames];append from rows; close
TranPMD;
FINISH TranPMD;

START TranPMC(parms,Age,Bage, APOE4);
   Q=J(4,4,0); rows={0 0 0 0 0 0 0 0 0 0 0 0 0};
   Q[1,2]=exp(parms[1]+parms[8]*bage+parms[15]*apoe4);
   Q[1,3]=exp(parms[2]+parms[9]*bage+parms[16]*apoe4);
   Q[1,4]=exp(parms[3]+parms[10]*bage+parms[17]*apoe4);
   Q[1,1]=-Q[1,2]-Q[1,3]-Q[1,4];
   Q[2,1]=exp(parms[4]+parms[11]*bage+parms[18]*apoe4);
```

```
   Q[2,3]=exp(parms[5]+parms[12]*bage+parms[19]*apoe4);
   Q[2,4]=exp(parms[6]+parms[13]*bage+parms[20]*apoe4);
   Q[2,2]=-Q[2,1]-Q[2,3]-Q[2,4];
   Q[3,4]=exp(parms[7]+parms[14]*bage+parms[21]*apoe4);
   Q[3,3]=-Q[3,4];
   A=teigvec(Q); V=teigval(Q); D=diag(exp(V[,1])); P=A*D*inv(A);
   do i=0 to Age;
      if i=0 then TPM=I(4); else TPM=TPM*P;
      row={0 0 0 0 0 0 0 0 0 0 0 0 0}; row[1]=Bage; row[2]=APOE4; row[3]=i;
      row[4]=TPM[1,1]; row[5]=TPM[1,2]; row[6]=TPM[1,3]; row[7]=TPM[1,4];
      row[8]=TPM[2,1]; row[9]=TPM[2,2]; row[10]=TPM[2,3]; row[11]=TPM[2,4];
      row[12]=TPM[3,3]; row[13]=TPM[3,4];rows=rows//row;
   end;
   varnames={"Bage" "APOE4" "Age" "P11C" "P12C" "P13C" "P14C" "P21C"
      "P22C" "P23C" "P24C" "P33C" "P34C"};
   create TranPMC from rows[COLNAME=varNames];append from rows; close
TranPMC;
FINISH TranPMC;

Data TranPData; merge tranpmd tranpmc; age=age+80; by age; where bage ne 0;
PROC SGplot data=TranPData noautolegend;
   SERIES x=age y=p13C/lineattrs=(color=black pattern=1 thickness=2);
   SERIES x=age y=p13D/lineattrs=(color=black pattern=2 thickness=2);
   xaxis label="Age" labelattrs=(size=16 weight=bold)
   valueattrs=(size=16 weight=bold) ;
   yaxis label="Probability" labelattrs=(size=16 weight=bold)
   valueattrs=(size=16 weight=bold) max=0.16;
Run;
PROC SGplot data=TranPData noautolegend;
   SERIES x=age y=p23C/lineattrs=(color=black pattern=1  thickness=2);
   SERIES x=age y=p23D/lineattrs=(color=black pattern=2  thickness=2);
   xaxis label="Age" labelattrs=(size=16 weight=bold)
   valueattrs=(size=16 weight=bold);
   yaxis label="Probability" labelattrs=(size=16 weight=bold)
   valueattrs=(size=16 weight=bold) values=(0 0.10 0.20 0.30);
Run;
%MEND;

/*****************************************************************************/
/*4 Goodness-of_fit: prevelance*/
/*****************************************************************************/
Data NunPrev; set Nun;
  if first_id then do; state=priorstate; vage=priorage-75; output; end;
  state=currentstate; vage=currentage-75; output;  keep id vage state bage lage death;
Run;
Data NunPrev; set Nunprev; by id vage; first_id=0; last_id=0;
  if first.id then first_id=1;
  if last.id then last_id=1; age=ceil(vage);
Run;
```

```
PROC IML;
START PrevObs;
   use nunprev;read all var{id age state first_id lage death} into dataset;close nunprev;
   n=nrow(dataset); obsv=J(617*40,7,0); Idn=0;
   do i=1 to n;
      id=dataset[i,1]; age=dataset[i,2]; state=dataset[i,3]; first_id=dataset[i,4];
      lage=dataset[i,5]; death=dataset[i,6];
      if death=0 then cutage=lage; else cutage=40;
      if first_id=1 then Idn=Idn+first_id; cstate=state;
      do j=1 to 40;
         obsv[(idn-1)*40+j, 1]=id; obsv[(idn-1)*40+j, 2]=j;
         if j>=age & j<=cutage then do;
            obsv[(idn-1)*40+j, 3]=cstate;
            if cstate=1 then obsv[(idn-1)*40+j, 4:7]={1 0 0 0};
            else if cstate=2 then obsv[(idn-1)*40+j, 4:7]={0 1 0 0};
            else if cstate=3 then obsv[(idn-1)*40+j, 4:7]={0 0 1 0};
            else if cstate=4 then obsv[(idn-1)*40+j, 4:7]={0 0 0 1};
         end;
      end;
   end;
   varnames={"ID" "Age" "state"  "state1"  "state2"  "state3"  "state4" };
   create PrevObs from obsv[colname=varNames];append from obsv; close PrevObs;
FINISH prevObs;
Run PrevObs;
Quit;

PROC SQL;
create table PrevObsP as select Age, sum(state1) as state1_OBS, sum(state2) as
state2_OBS,
sum(state3) as state3_OBS, sum(state4) as state4_OBS from PrevObs group by age;
Quit;
Data EpC; set Nun_4state; where first_id=1; keep id bstate bage apoe4 lage death;
RUN;

Proc IML;
START PrevEstC(parms);
   use EpC; read all var {ID bstate bage apoe4 lage death } into Dataset; close EpC;
   prevc=J(617*40,10,0);  Q=J(4,4,0);
   do i=1 to 617;
      id=dataset[i,1]; bstate=dataset[i,2]; bage=dataset[i,3];
      apoe4=dataset[i,4]; lage=dataset[i,5]; death=dataset[i,6];
      Q[1,2]=exp(parms[1]+parms[8]*bage+parms[15]*apoe4);
      Q[1,3]=exp(parms[2]+parms[9]*bage+parms[16]*apoe4);
      Q[1,4]=exp(parms[3]+parms[10]*bage+parms[17]*apoe4);
      Q[2,1]=exp(parms[4]+parms[11]*bage+parms[18]*apoe4);
      Q[2,3]=exp(parms[5]+parms[12]*bage+parms[19]*apoe4);
      Q[2,4]=exp(parms[6]+parms[13]*bage+parms[20]*apoe4);
      Q[3,4]=exp(parms[7]+parms[14]*bage+parms[21]*apoe4);
      Q[1,1]=-Q[1,2]-Q[1,3]-Q[1,4]; Q[2,2]=-Q[2,1]-Q[2,3]-Q[2,4]; Q[3,3]=-Q[3,4];
      A=teigvec(Q); V=teigval(Q);
```

```
      do age=1 to 40;
         prevc[(i-1)*40+age,1]=i; prevc[(i-1)*40+age,2]=age;
         prevc[(i-1)*40+age,3]=bstate;
         prevc[(i-1)*40+age,4]=bage; prevc[(i-1)*40+age,5]=lage;
         prevc[(i-1)*40+age,6]=death;
         if death=0 then cutage=lage; else cutage=40;
         if age>=bage & age<=cutage then do;
            D=diag(exp(V[,1]*(age-bage))); P=A*D*inv(A);
            prevc[(i-1)*40+age,7]=P[bstate,1]; prevc[(i-1)*40+age,8]=P[bstate,2];
            prevc[(i-1)*40+age,9]=P[bstate,3]; prevc[(i-1)*40+age,10]=P[bstate,4];
         end;
      end;
   end;
   varnames={"ID" "Age" "bstate" "bage" "lage" "death" "state1"  "state2"
     "state3"  "state4" };
   create Prevc from prevc[COLNAME=varNames];append from prevc; close prevc;
FINISH PrevEstC;

Proc IML;
START PrevEstD(parms);
   use EpC; read all var {ID bstate bage apoe4 lage death } into Dataset; close EpC;
   prevd=J(617*40,10,0); P=J(4,4,0);
   do i=1 to 617;
      id=dataset[i,1]; bstate=dataset[i,2]; bage=dataset[i,3];
      apoe4=dataset[i,4]; lage=dataset[i,5]; death=dataset[i,6];
      pexp11=1; pexp12=exp(parms[1]+parms[8]*bage+parms[15]*apoe4);
      pexp13=exp(parms[2]+parms[9]*bage+parms[16]*apoe4);
      pexp14=exp(parms[3]+parms[10]*bage+parms[17]*apoe4);
      pexp22=1; pexp21=exp(parms[4]+parms[11]*bage+parms[18]*apoe4);
      pexp23=exp(parms[5]+parms[12]*bage+parms[19]*apoe4);
      pexp24=exp(parms[6]+parms[13]*bage+parms[20]*apoe4);
      pexp33=1; pexp34=exp(parms[7]+parms[14]*bage+parms[21]*apoe4);
      pexp1=pexp11+pexp12+pexp13+pexp14;
      P[1,1]=pexp11/pexp1; P[1,2]=pexp12/pexp1;
      P[1,3]=pexp13/pexp1; P[1,4]=pexp14/pexp1;
      pexp2=pexp21+pexp22+pexp23+pexp24;
      P[2,1]=pexp21/pexp2; P[2,2]=pexp22/pexp2;
      P[2,3]=pexp23/pexp2; P[2,4]=pexp24/pexp2;
      P[3,3]=pexp33/(pexp33+pexp34); P[3,4]=pexp34/(pexp33+pexp34); P[4,4]=1;
      TPM=I(4);
      do age=1 to 40;
         TPM=TPM*P;
         prevd[(i-1)*40+age,1]=i; prevd[(i-1)*40+age,2]=age;
         prevd[(i-1)*40+age,3]=bstate;
         prevd[(i-1)*40+age,4]=bage; prevd[(i-1)*40+age,5]=lage;
         prevd[(i-1)*40+age,6]=death;
         if death=0 then cutage=lage; else cutage=40;
         if age>=bage & age<=cutage then do;
            prevD[(i-1)*40+age,7]=TPM[bstate,1];
            prevD[(i-1)*40+age,8]=TPM[bstate,2];
```

```
            prevD[(i-1)*40+age,9]=TPM[bstate,3];
            prevD[(i-1)*40+age,10]=TPM[bstate,4];
          end;
       end;
   end;
   varnames={"ID" "Age" "bstate" "bage" "lage" "death" "state1"  "state2"
    "state3"  "state4" };
   create PrevD from prevD[colname=varNames];append from prevD; close prevD;
FINISH PrevEstD;
Quit;
```

B. SAS codes for Non-homogenous Markov Model

```
/*************************************************************************/
/*1. %Macro SimData(SimN,SubjN); */
/*************************************************************************/
options set=R_HOME='C:\Program Files\R\R-2.15.2';
%Macro SimData(SimN,SubjN,betas);
%let rowN=%eval(&SimN * &SubjN);
PROC IML;
Nrows=&rowN; SN=&SimN; SbjN=&SubjN;
%include "~ \GenDataR.txt";
RUN ImportDataSetFromR("SimH&SubjN", "SimH");
RUN ImportDataSetFromR("SimG&SubjN", "SimG");
Quit;
%Mend SimData;

/*************************************************************************/
/*2. %Macro DataPrep(Dataset); */
/*************************************************************************/
%Macro DataPrep(Dataset);
Data SimD1; set &Dataset;
rename V1=SimN V2=Id V3=T1 V4=T2 V5=T3 V6=T4 V7=T5 V8=Z; Run;
Data SimD2; set SimD1;
   if T1<T2 and T1<T3 then do;State1=2;T=T1;TL1=floor(T);TR1=ceil(T);end;
   if T2<T1 and T2<T3 then do;State1=3;T=T2;TL1=floor(T);TR1=ceil(T);end;
   if T3<T1 and T3<T2 then do;State1=4;T=T3;TL1=floor(T);TR1=T;end;
   if T>25 then do; TL1=25; TR1=25; State1=1; end;
   if State1=2 then do;
        if T4<T5 then do; State2=3; TT=T+T4; TL2=floor(TT); TR2=ceil(TT);end;
        if T5<T4 then do; State2=4; TT=T+T5; TL2=floor(TT); TR2=TT;end;
        if TT>25 then do; State2=2; TL2=25;TR2=25; end;
   end;
   if TL1=TL2 then do;State1=State2; State2=.; TR1=TR2; TL2=.; TR2=.; end;
RUN;
Data SimD3;  set SimD2;
   if State1=1 then do;case=1; TL=0; TR=TR1;output;end;
   if State1=2 then do;
      case=1; TL=0; TR=TL1; output; case=2; TL=TL1; TR=TR1;output;
      if State2=2 then do;case=5; TL=TR1;TR=TR2;output; end;
```

```
      if State2=3 then do;
        case=5; TL=TR1;TR=TL2; output; case=6; TL=TL2;TR=TR2;output;
      end;
      if State2=4 then do;
        case=5; TL=TR1;TR=TL2; output; case=7; TL=TL2;TR=TR2;output;
      end;
   end;
   if State1=3 then do;
     case=1; TL=0; TR=TL1;output; case=3; TL=TL1;TR=TR1; output;
   end;
   if State1=4 then do;
     case=1; TL=0; TR=TL1;output; case=4; TL=TL1;TR=TR1; output;
   end;
  keep SimN id case TL TR Z;
Run;
Data SimD3; set SimD3; if TL ne TR; run;
Data C&Dataset; set SimD3; by SimN id;
   if first.id then first=1;else first=0; if last.id then last=1; else last=0;
   if TL=0 then TL=0.000001;
Run;
%Mend DataPrep;

/*************************************************************************/
/*3. %Macro WLLike;*/
/*************************************************************************/
%MACRO WLLike;
START logLikeWeibull(parms) global(Dataset);
   lamda12=parms[1];lamda13=parms[2];lamda14=parms[3];int12=parms[4];
   int13=parms[5]; int14=parms[6]; int23=parms[7];int24=parms[8];
   Z12=parms[9];Z13=parms[10];Z14=parms[11];
   logLike=.0; n=nrow(DataSet);
   do i=1 to n;
      case=DataSet[i,1];TL=DataSet[i,2];TR=DataSet[i,3];Z=DataSet[i,4];
      first=DataSet[i,5];last=DataSet[i,6];
      if first=1 then do;
         expZ12=exp(int12+Z12*Z);expZ13=exp(int13+Z13*Z);
         expZ14=exp(int14+Z14*Z); a23=exp(int23); a24=exp(int24);
      end;
      /*Case 1: 1->1*/
      if case=1 then  do;
         A12TL=expZ12*(TL)**lamda12; A13TL=expZ13*(TL)**lamda13;
         A14TL=expZ14*(TL)**lamda14;
         A12TR=expZ12*(TR)**lamda12; A13TR=expZ13*(TR)**lamda13;
         A14TR=expZ14*(TR)**lamda14;
         logL=(A12TL+A13TL+A14TL)-(A12TR+A13TR+A14TR);
      end;
      /*Case 2: 1->2*/
      if case=2 then do;
         intg=.0; h=(TR-TL)/150;
         do j=0 to 150;u=TL+j*h;
```

```
      *1. Get P11(TL,u);
      A12TL=expZ12*(TL)**lamda12; A13TL=expZ13*(TL)**lamda13;
      A14TL=expZ14*(TL)**lamda14;
      A12u=expZ12*(u)**lamda12; A13u=expZ13*(u)**lamda13;
      A14u=expZ14*(u)**lamda14;
      p11=exp((A12TL+A13TL+A14TL)-(A12u+A13u+A14u));
      *2. Get a12(u);
      q12=expZ12*(lamda12)*(u)**(lamda12-1);
      *3. Get p22(u,TR);
      p22=exp(-(a23+a24)*(TR-u));
      *4. do intergration p11(TL,u)*a12(u)*p22(u,TR);
      f=p11*q12*p22; if j=0 | j=150 then intg=intg+f/2; else intg=intg+f;
  end;
  logL=log(intg*h);
end;
/*Case 3: 1->3*/
if case=3 then do;
    intg=.0; h=(TR-TL)/150;
    do j=0 to 150; u=TL+j*h;
       *1. Get P11(TL,u);
       A12TL=expZ12*(TL)**lamda12; A13TL=expZ13*(TL)**lamda13;
       A14TL=expZ14*(TL)**lamda14;
       A12u=expZ12*(u)**lamda12; A13u=expZ13*(u)**lamda13;
       A14u=expZ14*(u)**lamda14;
       p11=exp((A12TL+A13TL+A14TL)-(A12u+A13u+A14u));
       *2. Get a12(u) a13(u);
       q12=expZ12*(lamda12)*(u)**(lamda12-1);
       q13=expZ13*(lamda13)*(u)**(lamda13-1);
       *3. Get p23(u,TR);
       p23= a23/(a23+a24)*(1-exp((a23+a24)*(u-TR)));
       *4. do intergration p11(TL,u)*(a13(u)+a12(u)*p23(u,TR));
       f=p11*(q12*p23+q13); if j=0 | j=150 then intg=intg+f/2; else intg=intg+f;
    end;
    logL=log(intg*h);
end;
/*Case 4: 1->4*/
if case=4 then do;
    intg=.0; h=(TR-TL)/150;
    *1. get p14(TL,TR)=p11(TL,TR)*a14(TR);
    A12TL=expZ12*(TL)**lamda12; A13TL=expZ13*(TL)**lamda13;
    A14TL=expZ14*(TL)**lamda14; A12TR=expZ12*(TR)**lamda12;
    A13TR=expZ13*(TR)**lamda13; A14TR=expZ14*(TR)**lamda14;
    p11TR=exp((A12TL+A13TL+A14TL)-(A12TR+A13TR+A14TR));
    q14TR=expZ14*(lamda14)*(TR)**(lamda14-1); p14=p11TR*q14TR;
    do j=0 to 150;
       u=TL+j*h;
       *1. Get P11(TL,u);
       A12TL=expZ12*(TL)**lamda12; A13TL=expZ13*(TL)**lamda13;
       A14TL=expZ14*(TL)**lamda14; A12u=expZ12*(u)**lamda12;
       A13u=expZ13*(u)**lamda13;A14u=expZ14*(u)**lamda14;
```

```
            p11=exp((A12TL+A13TL+A14TL)-(A12u+A13u+A14u));
            *2. Get a12(u);
            q12=expZ12*(lamda12)*(u)**(lamda12-1);
            *3. Get p24(u,TR);
            p22=exp(-(a23+a24)*(TR-u));p24=p22*a24;
            *4. do intergration p11(TL,u)*a12(u)*p24(u,TR));
            f=p11*q12*p24; if j=0 | j=150 then intg=intg+f/2; else intg=intg+f;
          end;
          logL=log(p14+intg*h);
        end;
        /*Case 5: 2->2*/
        if case=5 then logL=-(a23+a24)*(TR-TL);
        /*Case 6: 2->3*/
        if case=6 then do;
           p23=(1-exp(-(a23+a24)*(TR-TL)))*a23/(a23+a24); logL=log(p23);
        end;
        /*Case 7: 2->4*/
        if case=7 then  logL=-(a23+a24)*(TR-TL)+log(a24);
        logLike=logLike+logL;
     end;
     return(logLike);
   FINISH logLikeWeibull;
%Mend WLLike;

/*************************************************************************/
/*4. %Macro EstLJ(Dataset,Nsim,H0,Con); */
/*************************************************************************/
%Macro EstLJ(Dataset,Nsim,H0,Con);
  PROC IML;
   %WLLike; &H0; &Con; optn={1 0 1 3}; ct={1000 1000}; Est=J(&Nsim,55,0);
   use &Dataset;
   do sim=1 to &Nsim;
      read all var {Case TL TR Z First Last} into Dataset where(SimN=sim);
      call nlpnra(rc,xres,"logLikeWeibull",h0,optn,con,ct); estimate=xres` ;
      call nlpfdd(f,g,hes2,"logLikeWeibull",estimate);cov=-inv(hes2);
      norqua=probit(1-0.05/2); stderr=sqrt(vecdiag(cov));
      low=estimate-norqua*stderr; up=estimate+norqua*stderr;
      z=abs(estimate/stderr);p=2*(1-probnorm(z));
      Est[sim,1:11]=xres;Est[sim,12:22]=low`;Est[sim,23:33]=up`;
      Est[sim,34:44]=p`;Est[sim,45:55]=stderr`;
   end;
   varNames={"lamda12" "lamda13" "lamda14" "int12" "int13" "int14" "int23"
   "int24" "Z12" "Z13" "Z14" "lamda12CL" "lamda13CL" "lamda14CL" "int12CL"
   "int13CL" "int14CL" "int23CL" "int24CL" "Z12CL" "Z13CL" "Z14CL"
   "lamda12CU" "lamda13CU" "lamda14CU" "int12CU"
   "int13CU" "int14CU" "int23CU" "int24CU" "Z12CU" "Z13CU" "Z14CU"
   "lamda12p" "lamda13p" "lamda14p" "int12p" "int13p" "int14p" "int23p" "int24p"
   "Z12p" "Z13p" "Z14p" "lamda12SD" "lamda13SD" "lamda14SD" "int12SD"
   "int13SD" "int14SD" "int23SD" "int24SD" "Z12SD" "Z13SD" "Z14SD"};
```

```
   create R&Dataset from Est[COLNAME=varNames] ;Append from Est; close
R&Dataset;
Quit;
%Mend;

/*************************************************************************/
/*5. %Macro SimLJTable(Dataset);*/
/*************************************************************************/
%Macro SimLJTable(Dataset);
Data SimT; Set &Dataset (keep=lamda13 Z12 Z13 Z14 Z12CL Z13CL Z14CL
Z12CU Z13CU Z14CU Z12SD Z13SD Z14SD);Beta1=2; Beta2=1.5; Beta3=0;
Ebeta1=Z12;Ebeta2=Z13;
Ebeta3=Z14; Bias1=Ebeta1-Beta1;Bias2=Ebeta2-Beta2;Bias3=Ebeta3-Beta3;
SE1=Bias1**2; SE2=Bias2**2;SE3=Bias3**2;
if Z12CL<=2 and Z12CU>=2 then Cbeta1=1; else Cbeta1=0;
if Z13CL<=1.5 and Z13CU>=1.5 then Cbeta2=1; else Cbeta2=0;
if Z14CL<=0 and Z14CU>=0 then Cbeta3=1; else Cbeta3=0;
where lamda13 ne 0.01; Run;
Proc tabulate data=SimT; var Z12 Z13 Z14 Z12SD Z13SD Z14SD Se1 Se2 Se3 Bias1
Bias2 Bias3 Cbeta1 Cbeta2 Cbeta3;
table (Z12 Z13 Z14)*(N mean*f=8.4 std*f=8.4) (Se1 Se2 Se3)*Mean*f=8.4
(Z12SD Z13SD Z14SD)*Mean*f=8.4 (Bias1 Bias2 Bias3)*Mean*f=8.4
(Cbeta1 Cbeta2 Cbeta3)*Mean*f=percentn10.2; Run;
%Mend;

/*************************************************************************/
/*6. %Macro SimModel1LJ(Nsim,SubjN);*/
/*************************************************************************/
%Macro SimModel1LJ(Nsim,SubjN);
 %SimData(&Nsim,&SubjN);
  /*1. Homogenous Data */
 %DataPrep(SimH&SubjN);
 %let h0H=%str(h0={1 1 1 -4.2 -4.3 -4 -1.5 -2  2 1.5 0};);
 %let conH=%str(con={0.01 0.01  0.01  . . . . .  . . ., . . .  . . . . .  . . .};);
 %EstLJ(CSimH&SubjN,&Nsim,&H0H,&ConH); %SimLJTable(RCSimH&SubjN);
  /*2. Weibull Data  */
 %DataPrep(SimW&SubjN);
 %let H0W=%str(h0={1.9 2 2.1 -6.5 -6.7 -6 -1.5 -2  2 1.5 0};);
 %let ConW=%str(con={1.00001 1.00001  1.00001  . . . . .  . . ., . . .  . . . . .  . . .};);
 %EstLJ(CSimW&SubjN,&Nsim,&H0W,&ConW);
 %SimLJTable(RCSimW&SubjN);
  /*3. Gomperz Data  */
 %DataPrep(SimG&SubjN);
 %let H0G=%str(h0={3.5 5 5.5 -12.5 -17 -18 -1.5 -2  2 1.5 0};);
 %let ConG=%str(con={1.00001 1.00001  1.00001  . . . . .  . . .,  . . .  . . . . .  . . .};);
 %EstLJ(CSimG&SubjN,&Nsim,&H0G,&ConG);
%Mend SimModel1LJ;
```

C. R codes for Semi-Markov model

```
library(rngWELL)
library(randtoolbox)
a12=function(t,pars) {return(pars[2]*pars[1]*(t^(pars[1]-1)))}
a13=function(t,pars) {return(pars[4]*pars[3]*(t^(pars[3]-1)))}
a14=function(t,pars) {return(pars[6]*pars[5]*(t^(pars[5]-1)))}
a23<-function(u1,t,cutage,pars) {
 if (t>u1) {
  if (t>cutage) res=pars[8]*pars[7]*(t-u1)^(pars[7]-1)*exp(pars[9])
  else res=pars[8]*pars[7]*(t-u1)^(pars[7]-1)
 }
 else res=0
 return(res)
}

a24<-function(u1,t,cutage,pars) {
 if (t>u1) {
  if (t>cutage) res=pars[11]*pars[10]*(t-u1)^(pars[10]-1)*exp(pars[12])
  else res=pars[11]*pars[10]*(t-u1)^(pars[10]-1)
 }
 else res=0
 return(res)
}
p11=function(t,pars)
 return(exp(-pars[2]*(t^pars[1])-pars[4]*(t^pars[3])-pars[6]*(t^pars[5])))
p22=function(u1,u2,cutage,pars){
 if (cutage>u2) A2=pars[8]*(u2-u1)^pars[7]+pars[11]*(u2-u1)^pars[10]
 else if (cutage>u1) A2=pars[8]*exp(pars[9])*(u2-u1)^pars[7]
   +pars[8]*(1-exp(pars[9]))*(cutage-u1)^pars[7]+ pars[11]*exp(pars[12])
   *(u2-u1)^pars[10]+pars[11]*(1-exp(pars[12]))*(cutage-u1)^pars[10]
 else A2=pars[8]*exp(pars[9])*(u2-u1)^pars[7]+pars[11]*exp(pars[12])
   *(u2-u1)^pars[10]
 return(exp(-A2))
}

#case 2: 1->2 or 1->2->2
p12<-function(t1,t2,t5,cutage,pars,SN){
 pf12=function(u,t5,cutage,pars)
  return(p11(u,pars)*a12(u,pars)*p22(u,t5,cutage,pars))
 r1<-t1+(t2-t1)*halton(2*SN)
 #r1<-t1+(t2-t1)*sobol(2*SN,scrambling=1)
 res<-rep(0,SN)
 for (i in 1:SN) res[i]=pf12(r1[i+SN],t5,cutage,pars)
 return(mean(res)*(t2-t1))
}

#case 3: 1->3
p13<-function(t1,t2,cutage,pars,SN){
 pf13<-function(u,pars) return(p11(u,pars)*a13(u,pars))
```

```
  pf123<-function(u1,u2,cutage,pars) {
    p22e=ifelse(u2>u1,p22(u1,u2,cutage,pars)*a23(u1,u2,cutage,pars),0)
    res=p11(u1,pars)*a12(u1,pars)*p22e
    return(res)}
  r1<-t1+(t2-t1)*halton(2*SN)
  r2<-t1+(t2-t1)*halton(n=2*SN,dim=2)
  #r1<-t1+(t2-t1)*sobol(2*SN,,scrambling=1)
  #r2<-t1+(t2-t1)*sobol(n=2*SN,dim=2,,scrambling=1)
  res1<-rep(0,SN)
  res2<-rep(0,SN)
  for (i in 1:SN) {
    res1<-pf13(r1[i+SN],pars)
    res2<-pf123(r2[i+SN,1],r2[i+SN,2],cutage,pars)
  }
  p1300=mean(res1)*(t2-t1)
  p1230<-mean(res2)*(t2-t1)^2
  return(p1300+p1230)
}

#case 4: 1->4
p14<-function(t,pars) return(p11(t,pars)*a14(t,pars))

#Case 5: 1->2->3
p123<-function(t1,t2,t3,t4,cutage,pars,SN){
  pf123<-function(u1,u2,cutage,pars)
     return(p11(u1,pars)*a12(u1,pars)*p22(u1,u2,cutage,pars)*a23(u1,u2,cutage,pars))
  r1<-halton(2*SN,dim=2)
  #r1<-sobol(2*SN,dim=2,,scrambling=1)
  r1[,1]=t1+(t2-t1)*r1[,1]
  r1[,2]=t3+(t4-t3)*r1[,2]
  res<-rep(0,SN)
  for (i in 1:SN) res[i]<-pf123(r1[i+SN,1],r1[i+SN,2],cutage,pars)
  return(mean(res)*(t2-t1)*(t4-t3))
}

#Case 6: 1->2->4
p124<-function(t1,t2,t5,cutage,pars,SN){
  pf124<-function(u,t5,cutage,pars)
     return(p11(u,pars)*a12(u,pars)*p22(u,t5,cutage,pars)*a24(u,t5,cutage,pars))
  r1<-t1+(t2-t1)*halton(2*SN)
  #r1<-t1+(t2-t1)*sobol(2*SN,,scrambling=1)
  res<-rep(0,SN)
  for (i in 1:SN) res[i]<-pf124(r1[i+SN],t5,cutage,pars)
  return(mean(res)*(t2-t1))
}
# Main log-likelihood function
logLSemi<-function(parms,dataset,cutage,SN) {
  assign("dataset",dataset,envir=.GlobalEnv)
  k12=parms[1];k13=parms[2];k14=parms[3]; k23=parms[4];k24=parms[5]
  int12=parms[6]; int13=parms[7];int14=parms[8];int23=parms[9];int24=parms[10]
```

```
 par23_t=0;
 par24_t=parms[11]
 Apoe12=parms[12]; Female14=parms[13]
 Sm14=parms[14]; Sm23=parms[15]
 n=nrow(dataset);logLike=.0
 for (i in 1:n){
  case=dataset[i,1]
  Bage=dataset[i,2]; t1=dataset[i,3]; t2=dataset[i,4]; t3=dataset[i,5]; t4=dataset[i,6];
  t5=dataset[i,7];
  Apoe=dataset[i,8];Female=dataset[i,9]; Educ=dataset[i,10]; Fam=dataset[i,11]
  Db=dataset[i,12]; Sm=dataset[i,13];Hd=dataset[i,14]
  expz12=exp(int12+Apoe12*Apoe)
  expz13=exp(int13)
  expz14=exp(int14+Female14*Female+Sm14*Sm)
  expz23=exp(int23+Sm23*Sm)
  expz24=exp(int24)
  pars=c(k12,expz12,k13,expz13,k14,expz14,k23,expz23,par23_t,k24,
    expz24,par24_t)
  if (case==1) Li=p11(t5,pars)
  else if (case==2) Li=p12(t1,t2,t5,cutage,pars,SN)
  else if (case==3) Li=p13(t1,t2,cutage,pars,SN)
  else if (case==4) Li=p14(t5,pars)
  else if (case==5) Li=p123(t1,t2,t3,t4,cutage,pars,SN)
  else Li=p124(t1,t2,t5,cutage,pars,SN)
  logLike=logLike+log(Li)-log(p11(Bage,pars))
 }
 return(-logLike)
}
```

Bibliography

1. Commenges D. Inference for multi-state models from interval-censored data. *Stat Methods Med Res* 2002; **11**: 167-182.
2. Abner EL, Nelson PT, Schmitt FA, Browning SR, Fardo DW, Wan LJ, Jicha GA, Cooper GE, Smith CD, Caban-Holt AM, Van Eldik LJ, Kryscio RJ. Self-Reported Head Injury and Risk of Late-Life Impairment and AD Pathology in an AD Center Cohort. *Dement Geriatr Cogn Disord* 2014; **37**: 294-306.
3. Andersen PK, Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res* 2002; **11**: 91-115.
4. Meira-Machado L, de Una-Alvarez J, Cadarso-Suarez C, Andersen PK. Multi-state models for the analysis of time-to-event data. *Stat Methods Med Res* 2009; **18**: 195-222.
5. Commenges D. Multi-state models in epidemiology. *Lifetime Data Anal* 1999; **5**: 315-327.
6. Hougaard P. Multi-state models: a review. *Lifetime Data Anal* 1999; **5**: 239-264.
7. Hsieh HJ, Chen THH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Stat Med* 2002; **21**: 3369-3382.
8. van den Hout A, Matthews FE. Multi-state analysis of cognitive ability data: a piecewise-constant model and a Weibull model. *Stat Med* 2008; **27**: 5440-5455.
9. Frydman H. A Nonparametric-Estimation Procedure for a Periodically Observed 3-State Markov Process, with Application to Aids. *Journal of the Royal Statistical Society Series B-Methodological* 1992; **54**: 853-866.
10. Frydman H, Szarek M. Nonparametric Estimation in a Markov "Illness-Death" Process from Interval Censored Observations with Missing Intermediate Transition Status. *Biometrics* 2009; **65**: 143-151.
11. Joly P, Commenges D, Letenneur L. A penalized likelihood approach for arbitrarily censored and truncated data: Application to age-specific incidence of dementia. *Biometrics* 1998; **54**: 185-194.
12. Joly P, Commenges D. A penalized likelihood approach for a progressive three-state model with censored and truncated data: Application to AIDS. *Biometrics* 1999; **55**: 887-890.
13. Joly P, Commenges D, Helmer C, Letenneur L. A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia. *Biostatistics* 2002; **3**: 433-443.
14. Satten GA, Sternberg MR. Fitting semi-Markov models to interval-censored data with unknown initiation times. *Biometrics* 1999; **55**: 507-513.
15. Foucher Y, Mathieu E, Saint-Pierre P, Durand JF, Daures JP. A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal* 2005; **47**: 825-833.
16. Foucher Y, Giral M, Soulillou JP, Daures JP. A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. *Stat Methods Med Res* 2010; **19**: 127-145.
17. Kapetanakis V, Matthews FE, Hout A. A semi-Markov model for stroke with piecewise-constant hazards in the presence of left, right and interval censoring. *Stat Med* 2013; **32**: 697-713.

18. Salazar JC, Schmitt FA, Yu L, Mendiondo MM, Kryscio RJ. Shared random effects analysis of multi-state Markov models: application to a longitudinal study of transitions to dementia. *Stat Med* 2007; **26**: 568-580.
19. Yu L, Tyas SL, Snowdon DA, Kryscio RJ. Effects of ignoring baseline on modeling transitions from intact cognition to dementia. *Computational Statistics & Data Analysis* 2009; **53**: 3334-3343.
20. Abner EL, Kryscio RJ, Cooper GE, Fardo DW, Jicha GA, Mendiondo MS, Nelson PT, Smith CD, Van Eldik LJ, Wan L. Mild cognitive impairment: statistical models of transition using longitudinal clinical data. *International Journal of Alzheimer's Disease* 2012; **2012**.
21. Kryscio RJ, Abner EL, Cooper GE, Fardo DW, Jicha GA, Nelson PT, Smith CD, Van Eldik LJ, Wan L, Schmitt FA. Self-reported memory complaints Implications from a longitudinal cohort with autopsies. *Neurology* 2014; **83**: 1359-1365.
22. Kryscio RJ, Abner EL, Lin YS, Cooper GE, Fardo DW, Jicha GA, Nelson PT, Smith CD, Van Eldik LJ, Wan LJ, Schmitt FA. Adjusting for Mortality when Identifying Risk Factors for Transitions to Mild Cognitive Impairment and Dementia. *Journal of Alzheimers Disease* 2013; **35**: 823-832.
23. Agresti A. *Categorical data analysis.* (2nd edn). Wiley-Interscience: New York, 2002.
24. Kryscio RJ, Schmitt FA, Salazar JC. Risk factors for transitions from normal to mild cognitive impairment and dementia. *Neurology* 2006; **66**: 828-832.
25. Andersen AH, Smith CD, Slevin JT, Kryscio RJ, Martin CA, Schmitt FA, Blonder LX. Dopaminergic Modulation of Medial Prefrontal Cortex Deactivation in Parkinson Depression. *Parkinsons Disease* 2015.
26. Andersen PK, Esbjerg S, Sorensen TIA. Multi-state models for bleeding episodes and mortality in liver cirrhosis. *Stat Med* 2000; **19**: 587-599.
27. Hubbard RA, Zhou XH. A comparison of non-homogeneous Markov regression models with application to Alzheimer's disease progression. *Journal of Applied Statistics* 2011; **38**: 2313-2326.
28. Inc S. SAS/STAT® 9.3 User's Guide. *Cary, North Carolina: SAS Institute Inc* 2011.
29. Jackson CH. Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software* 2011; **38**: 1-28.
30. Institute S. *SAS/IML 9.3 User's Guide.* SAS Institute, 2011.
31. Titman AC, Sharples LD. A general goodness-of-fit test for Markov and hidden Markov models. *Stat Med* 2008; **27**: 2177-2195.
32. Abner EL, Charnigo RJ, Kryscio RJ. Markov chains and semi-Markov models in time-to-event analysis. *Journal of biometrics & biostatistics* 2013: 19522.
33. Kalbfleisch JD, Lawless JF. The Analysis of Panel Data under a Markov Assumption. *Journal of the American Statistical Association* 1985; **80**: 863-871.
34. Gentleman RC, Lawless JF, Lindsey JC, Yan P. Multistate Markov-Models for Analyzing Incomplete Disease History Data with Illustrations for Hiv Disease. *Stat Med* 1994; **13**: 805-821.
35. Hsieh HJ, Chen TH, Chang SH. Assessing chronic disease progression using non-homogeneous exponential regression Markov models: an illustration using a selective breast cancer screening in Taiwan. *Stat Med* 2002; **21**: 3369-3382.
36. Schmitt FA, Nelson PT, Abner E, Scheff S, Jicha GA, Smith C, Cooper G, Mendiondo M, Danner DD, Van Eldik LJ, Caban-Holt A, Lovell MA, Kryscio RJ. University of Kentucky Sanders-Brown Healthy Brain Aging Volunteers: Donor

Characteristics, Procedures and Neuropathology. *Current Alzheimer Research* 2012; **9**: 724-733.
37. David HA, Moeschberger ML. *The theory of competing risks.* Griffin London, 1978.
38. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. *Stat Med* 2009; **28**: 956-971.
39. Andersen PK. Multistate Models in Survival Analysis - a Study of Nephropathy and Mortality in Diabetes. *Stat Med* 1988; **7**: 661-670.
40. Mathieu E, Foucher Y, Dellamonica P, Daures JP. Parametric and non homogeneous semi-markov process for HIV control. *Methodology and Computing in Applied Probability* 2007; **9**: 389-397.
41. Putter H, van der Hage J, de Bock GH, Elgalta R, van de Velde CJH. Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal* 2006; **48**: 366-380.
42. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical models based on counting processes.* Springer Science & Business Media, 2012.
43. Titman AC, Sharples LD. Model diagnostics for multi-state models. *Stat Methods Med Res* 2010; **19**: 621-651.
44. Bhat CR. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transportation Research Part B-Methodological* 2001; **35**: 677-693.
45. Wei S, Kryscio RJ. Semi-Markov models for interval censored transient cognitive states with back transitions and a competing risk. *Stat Methods Med Res* 2014: 0962280214534412.

Vita

Lijie Wan

Education

M.S. in Statistics, University of Kentucky, 2010-2012

Employment

Teaching Assistant, August 2010-May 2011

Department of Statistics, University of Kentucky

Research Assistant, August 2011-March 2016

Sanders-Brown Center on Aging, University of Kentucky

Publications

Abner, E.L., Schmitt, F.A., Nelson, P.T., Lou, W., **Wan, L**., Gauriglia, R., Dodge, H.H., Woltjer, R.L., Yu, L., Bennet, D.A. and Schneider, JA. The Statistical Modeling of Aging and Risk of Transition Project: Data collection and harmonization across 11 longitudinal cohort studies of aging, cognition, and dementia. Observational studies 1.2015 (2015): 56.

Chuan-hua Wei, **Lijie Wan**, Chunling Liu (2014). Efficient Estimation in Heteroscedastic Partially Linear Varying Coefficient Models. Communications in Statistics-Simulation and Computation 44.4 (2015): 892-901.

Richard J. Kryscio, Erin L. Abner, Gregory E. Cooper, David W. Fardo, Gregory A. Jicha, Peter T. Nelson, Charles D. Smith, Linda J. Van Eldik, **Lijie Wa**n and Frederick A. Schmitt. Self-reported memory complaints Implications from a longitudinal cohort with autopsies. Neurology 83.15 (2014): 1359-1365.

Erin L Abner, Peter T Nelson, Frederick A Schmitt, Steven R Browning, David W Fardo, **Lijie Wan**, Gregory A Jicha, Gregory E Cooper, Charles D Smith, Allison M Caban-Holt, Linda J Van Eldik, Richard J Kryscio (2013). Self-Reported Head Injury and Risk of Late-Life Impairment and AD Pathology in an AD Center Cohort. Dementia and geriatric cognitive disorders. Vol. 37, No. 5-6, 2014.

Richard J Kryscio, Erin L Abner, Yushun Lin, Gregory E Cooper, David W Fardo, Gregory A Jicha, Peter T Nelson, Charles D Smith, Linda J Van Eldik, **Lijie Wa**n, Frederick A Schmitt (2013). Adjusting for Mortality when Identifying Risk Factors for Transitions to Mild Cognitive Impairment and Dementia. Journal of Alzheimer's Disease 35.4 (2013): 823-832.

Erin L Abner, Richard J Kryscio, Gregory E Cooper, David W Fardo, Gregory A Jicha, Marta S Mendiondo, Peter T Nelson, Charles D Smith, Linda J Van Eldik, **Lijie Wan**, Frederick A Schmitt. Mild cognitive impairment: statistical models of transition using longitudinal clinical data. International Journal of Alzheimer’s Disease 2012 (2012).