# Statistical Methods for Event History Data under Response Dependent Sampling and Incomplete Observation

by

Yidan Shi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2020

**Examining Committee Membership**

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Yingwei (Paul) Peng
Professor, Department of Public Health Science
Department of Mathematics and Statistics
Queen's University

Supervisor(s): Leilei Zeng
Associate Professor, Department of Statistics and Actuarial Science
University of Waterloo
Mary E. Thompson
Professor Emerita, Department of Statistics and Actuarial Science
University of Waterloo

Internal Member(s): Joel A. Dubin
Associate Professor, Department of Statistics and Actuarial Science
School of Public Health and Health Systems
University of Waterloo
Peisong Han
Assistant Professor, School of Public Health
University of Michigan

Internal-External Member: Suzanne L. Tyas
Associate Professor, School of Public Health and Health Systems
University of Waterloo

## Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

**Abstract**

This thesis discusses statistical problems in event history data analysis including survival analysis and multistate models. Research questions in this thesis are motivated by the Nun Study, which contains longevity data and longitudinal follow-up of cognition functions in 678 religious sisters. Our research interests lie in modeling the survival pattern and the disease process for dementia. These data are subject to a process-dependent sampling scheme, and the homogeneous Markov assumption is violated when using a multistate model to fit the panel data for cognition. In this thesis, we formulated three statistical questions according to the aforementioned issues and propose approaches to deal with these problems.

Survival analysis is often subject to left-truncation when the data are collected within certain study windows. Naive methods ignoring the sampling conditions yield invalid estimates. Much work has been done to deal with the bias caused by left-truncation. However, discussion on the loss-in-efficiency is limited. In Chapter 2, we proposed a method in which auxiliary information is borrowed to improve the efficiency in estimation. The auxiliary information includes summary-level statistics from a previous study on the same cohort and census data for a comparable population. The likelihood and score functions are developed. A Monte Carlo approximation is proposed to deal with the difficulty in obtaining tractable forms of the score and information functions. The method is illustrated by both simulation and real data application to the Nun Study.

Continuous-time Markov models are widely used for analyzing longitudinal data on the disease progression over time due to the great convenience for computing the probability transition matrices and the likelihood functions. However, in practice, the Markov assumption does not always hold. Most of the existing methods relax the Markov assumption while losing the advantage of that assumption in the calculation of transition probabilities. In Chapter 3, we consider the case where the violation of the Markov property is due to multiple underlying types of disease. We propose a mixture hidden Markov model where the underlying process is characterized by a mixture of multiple time-homogeneous Markov chains, one for each disease type, while the observation process contains states corresponding to the common symptomatic stages of these diseases. The method can be applied

to modeling the disease process of Alzheimer's disease and other types of dementia. In the Nun Study, autopsies were conducted on some of the deceased participants so that one can know whether these individuals have Alzheimer's pathology in their brains. Our method can incorporate these partially observed pathology data as disease type indicators to improve the efficiency in estimation. The predictions for the overall prevalence and type-specific prevalence for dementia are calculated based on the proposed method. The performance of the proposed methods is also evaluated via simulation studies.

Many prospective cohort studies of chronic diseases select individuals whose observed process history satisfies particular conditions. For instance, studies aiming to estimate the incidence rate of dementia or the effect of genetic factors on the disease would recruit individuals in the condition of being alive and disease-free. In contrast, some other studies may aim to collect information on disease progression or mortality from the time of the disease onset. Under such settings, individuals are recruited if they are in a subset of the states at the study entry, and the methods of estimation need to account for such state-dependent selection conditions. For multistate analysis, one option is to construct the likelihood based on the prospective data given the history up to and including the time at accrual. This approach yields consistent estimates under state-dependent sampling condition with a price of loss in efficiency. Alternatively, the likelihood contribution from the retrospective and current status data at the time of accrual can be incorporated, but with difficulty in obtaining such information. For example, subjects' initial states are often unknown, imposing a challenge for the computation of the contribution from the current status data at the time of recruitment. However, auxiliary information on the initial states may be available, such as the age-specific population prevalence data related to the disease. In Chapter 4, we proposed a weighted-likelihood method to incorporate auxiliary prevalence data and account for the state-dependent selection condition. The method is demonstrated by simulation and applied to the Nun Study of aging and Alzheimer's disease. A Bayesian sensitivity test is conducted to evaluate the impact of misspecification of the auxiliary prevalence.

**Acknowledgements**

I would like to express my sincerest gratitude to my supervisors Dr. Leilei Zeng and Dr. Mary E. Thompson for providing invaluable insight, guidance and support throughout my thesis' research. I also truly appreciate Leilei and Mary for always being patient, caring, and encouraging even during my most struggling time. I am the luckiest to have worked with them.

I would like to thank Dr. Joel A. Dubin, Dr. Peisong Han, Dr. Yingwei (Paul) Peng and Dr. Suzanne L. Tyas for providing inspiring questions and helpful comments on both statistical and epidemiological aspects. Special thanks go to Dr. Suzanne L. Tyas who has shared with me many experience and advice from her expertise in public health research, which has widen my knowledge a lot.

I want to take this opportunity to thank my master's supervisor, Dr. Duncan J. Murdoch for inspiring my research interest in applied statistics. I also thank Ms. Mary Lou Dufton and Mr. Greg Preston for their continuous help in administrative and technical issues; and Dr. Pengfei Li who made a course, which I was supposed to suffer, enjoyable.

Many thanks to my friends from the department Junhan, Meng, Mike, Shixiao, the group of my werewolf buddies and others who showed up in this memorable journey. Special thanks go to Danqiao, who I met on my first day in Waterloo. The story has became better with them.

Last, but not the least, I would like to express my heartfelt appreciation to my parents Lejuan Dong and Baodong Shi for their endless love, support and confidence in me.

## Dedication

To my parents and grandparents.

# Table of Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Overview

In this thesis, we will describe three research projects dealing with problems arising in event history data anaysis, including the left-truncation in survival analysis, process-dependent sampling in multistate models, and the violation of the homogeneous Markov property in multistate models. The connection between the three projects is that we are aiming to introduce auxiliary information to the estimation procedure. We will discuss the approaches to incorporate the auxiliary information and to deal with the uncertainty brought by the auxiliary information.

The specific circumstances include incorporation of partial summary level data to failure time data subject to left-truncation (Chapter 2); incorporation of partial knowledge of mixture component membership and parameter constraints for a non-Markov multistate model (Chapter 3); utilization of population prevalence information to adjust for a sampling condition in multistate analysis (Chapter 4).

### 1.1.1 Left-truncation in Failure Time Data

Survival analysis is a branch of statistical methods for failure time or time-to-event data. Examples include analysis of longevity in a population or modeling times to onset or relapse of specific diseases. These kinds of data are often subject to left-truncation. For instance, when considering the duration of survival with a disease, the data are typically collected from participants who had the disease and remained alive at the time of accrual. This implies that their survival time since disease onset has to be longer than the time to study entry, which leads to the so-called left-truncation of the survival times. Failure to consider this sampling condition will lead to invalid inference.

There has been considerable discussion on the estimation of a survival function under left-truncation. The proposed statistical methods can be generally classified as conditional or unconditional approaches (Asgharian et al., 2002). The former develops inference based on the likelihood conditional on the observed truncation times, for example conditional non-parametric estimation of survival functions for left-truncated data (e.g. Wang et al., 1986; Wang, 1987; Tsai et al., 1987), and conditional partial likelihood methods for the Cox models (e.g. Kalbfleisch and Lawless, 1991; Keiding and Moeschberger, 1992; Wang et al., 1993). Unconditional methods have been discussed by Vardi (1989), Wang (1996), Qin and Shen (2010), and Qin et al. (2011). The unconditional analysis may have an efficiency advantage over a conditional approach when it is possible to parameterize the distribution of the truncation times. For instance Wang (1991) pointed out a non-parametric maximum likelihood estimator (i.e. Vardi, 1989) can be more efficient than the conditional nonparametric product limit estimate when the truncation time follows an uniform distribution.

An alternative way to deal with loss in efficiency due to left-truncation, namely incorporation of auxiliary information from the same or similar populations, has drawn increasing attention in recent years. There are two papers (Li and Qin, 1998; Shen, 2014) proposing methods that incorporate information on the size of the “full" study population for estimation of a (conditional) survival function in a nonparametric setting. Li and Qin (1998) proposed a likelihood-based method by adding an extra component to the observed likelihood representing the contribution of the total number of truncated subjects. Shen (2014) proposed a constrained expectation-maximization (EM) algorithm for estimation.

In addition to the augmented likelihood approaches, Faucett et al. (2002) considered a method based on multiple imputation using auxiliary variables to deal with the loss in efficiency due to a large right-censored proportion.

In this thesis, we will describe a likelihood-based approach that makes use of auxiliary information from the truncated sample to improve the efficiency. The approach involves generating pseudo data and jointly modeling the pseudo sample with the left-truncated sample. A one-step Monte-Carlo expectation-maximization algorithm is developed for the inference based on an augmented likelihood. The gain in efficiency is measured by the relative reduction in variances for the estimates.

### 1.1.2 Non-Markov Multistate Models for Event History Data

Time-homogeneous Markov models are most widely used in multistate analysis for the ease of the derivation of the likelihood, score function and information function. However, this is a fairly strict assumption. Increasing attention has been brought to cases where the transition intensities are not constant over time. Methods using piecewise constant transition intensities provide a neat extension to the time-homogeneous Markov model, as the existing method can be applied within each piece. Discussions on such models appear in Gentleman et al. (1994), Andersen and Keiding (2002). Methods for semi-Markov models allow the transition intensities to depend on the sojourn time in each state. Satten and Sternberg (1999) proposed a semi-Markov model with unknown initiation times. Foucher et al. (2005) applied the semi-Markov model with generalized Weibull intensities to HIV disease process. And Foucher et al. (2007) applied a multiple terminal model to renal transplantation data. Kang and Lagakos (2006) proposed a semi-Markov model for panel data.

A Hidden Markov Model (HMM) is defined as a process where the observation model is conditioned on the status of an underlying Markov model. It can be used to deal with model misspecification for the observed data sequence. For example, while the observed illness-death process may not be Markovian, it might be reasonable to assume it is the case for an underlying process with several extra hidden states. Smyth (1994) proposed a

Hidden Markov based method for fault detection in continuous monitoring of a complex dynamic system. In addition, a HMM can be used to approximate a semi-Markov model due to the nice Markov property for the underlying process with hidden states. Titman (2014) proposed an approximation for semi-Markov model using Hidden Markov Models where the phase-type models are specified for the underlying process.

In this thesis, we will discuss a new hidden multistate model where the underlying process is characterized by a finite mixture of time-homogeneous Markov chains. Each chain process corresponds to one sub-type of the disease. Disease sub-types may be classified by biomarkers or medical examinations. Thus we hope to incorporate the known-membership in the estimating procedure so that the efficiency can be improved.

### 1.1.3 Process-dependent Sampling in Multistate Models

Event history studies are typically conducted by monitoring subjects over time and collecting information on occurrence of certain events according to a multistate dynamic disease process. The distinctive features include longitudinal observations and multiple events, and one obtains so-called panel data when the observation of the continuous-time multistate process is made at arbitrary time points. Multistate models are naturally suited and widely used for event history data analysis. When the model is assumed satisfying the Markov assumption, Kalbfleisch and Lawless (1985) proposed a likelihood-based method for panel data. There have been considerable research and applications (Andersen and Keiding, 2002; Tuma et al., 1979; Commenges, 2002) based on this work, as well as software packages such as the msm package in R (Jackson et al., 2011) developed for implementation.

Most of the multistate models and their extensions are discussed under an ideal situation with a randomly selected sample being followed up since the beginning of the process. However, in many longitudinal studies, for example prevalence cohort studies, the cohort is selected conditioning on being in certain stage of the disease process at the study entry. The information on the occurrence of events may be collected retrospectively and prospectively limited to a certain time window. This leads to truncation and censoring of event times. As in the case of survival analysis, lack of consideration of the process-dependent sampling

condition will lead to bias in multistate analysis. Joly and Commenges (1999) pointed out that left-truncation and right censoring are commonly arising problems in multistate models applied to epidemiology data. They proposed a penalized likelihood method for an illness-death model subject to both left-truncation and right censoring. Nonparametric likelihood-based approaches for panel data subject to left-truncation and right censoring were discussed in Frydman (1995) and Hudgens (2005).

Instead of constructing the likelihoods based on the prospective data conditioning on the process history up to and including time at which the subject is recruited (e.g. the duration from time origin to study entry and the states at entry time), we propose to use a weighted likelihood where the weights resemble the contribution of the current status data at the time of recruitment given the sampling condition. As the subjects' initial states are unknown, auxiliary information such as age-specific population prevalences are used to facilitate the calculation of the weights.

## 1.2 Motivation: The Nun Study

The Nun Study is a longitudinal study started in 1991 and ended in 2003. The participants were 678 Catholic sisters, whose cognitive functions and survival were followed over the study period. These participants are a sub-sample from a population of 3926 American Roman Catholic sisters from the School Sisters of Notre Dame who were born between 1886 and 1916. There have been multiple studies conducted on this birth cohort. The two studies discussed in this thesis include the Nun Study as already before, as well as a mortality study (Butler and Snowdon, 1996) conducted from 1965 to 1989.

The Nun Study involved 678 participants who were alive and at least 75 years old at the study entry which is 1991. The observations include yearly measurements of the cognitive functions for the participants until death or end of the study. Among the 678 participants, 676 have more than one follow-up observations, 297 of them were diagnosed with dementia, and 606 died during the follow-up period. Postmortem neuoropathologic assessments were conducted on the sisters who passed away during the study period. In addition, a variety of information was collected from participants including: (1) year of birth; (2) the presence

of the allele $\epsilon 4$ in the apolipoprotein E gene ($APOE - \epsilon 4$), which is a commonly known risk factor for Alzheimer's disease; (3) intellectual factors such as educational level, high school course grades, and number of languages spoken.

A Mortality Study was conducted on the same birth cohort of Catholic sisters (Butler and Snowdon, 1996). Participants were 2573 sisters who were alive and aged 50 and over at the study baseline 1965. The participants were followed until 85 years old or 1989 if they had not reached 85 by then. During the study, 1103 deaths were observed at ages between 50 and 84. Unfortunately, the individual level death times are no longer available. However, summarized data from this mortality study are reported in Butler and Snowdon (1996), including the total number of deaths, the standard mortality ratio (SMR) of the nun cohort versus the US white female population, and information on the birth time distribution of the original birth cohort.

In Chapter 2, the research interest lies on modeling the mortality of the population of Catholic sisters. In Chapter 3 and Chapter 4, we focus on modeling the dementia and Alzheimer's disease process, and evaluating risk factors using multistate models.

## 1.3 Outline

In Chapter 2, we propose a method where we incorporate the auxiliary information from the Mortality Study with the data from the Nun Study to improve the efficiency when estimating the survival distribution. Chapter 2 is organized as follows. First, we will describe the question and define the notations in Section 2.1. Then, in Section 2.2, we will develop a multiple imputation based method to incorporate the summary level information by creating pseudo samples and joint modeling. In Section 2.3, we will provide the details and results of application to the Nun Study data, as well as a sensitivity test of model misspecification for the auxiliary information. A simulation study is conducted and given in Section 2.4, which is aiming to verify the variances estimation and explore the gain in efficiency. Section 2.5 is a discussion about the benefits and drawbacks of our methods, as well as potential future work.

In Chapter 3, we propose a mixture hidden Markov model to incorporate the pathology information. In Section 3.1, we introduce the motivation. In Section 3.2, we introduce the finite mixture hidden Markov model (HMM) and a likelihood-based method for multiple disease types. This method accounts for the heterogeneity in the disease processes and panel observation on disease occurrence. Disease type information is partially available based on the pathology data. In Section 3.3, we apply the mixture HMM to the Nun Study. In Section 3.4, we conduct simulation studies to illustrate the performance of the proposed method. Then in Section 3.5, we give some discussion of the identifiability and estimability issues arising in HMMs. In Section 3.6, we use a Bayesian method as an alternative way to get the estimates and handle the identifiability issues. Section 3.7 discusses future work.

In Chapter 4, we consider a weighted likelihood method to deal with the process-dependent sampling condition in event history studies. We first explain the motivation question in Section 4.1. In Section 4.2, we define notations and introduce methods for multistate Markov models under intermittent observation. Then we develop a weighted likelihood which accommodates auxiliary population prevalence information to adjust for sampling condition. In Section 4.3, we apply the weighted likelihood to study the onset of dementia and effects of important factors using the Nun Study data. In Section 4.4, we conduct simulation studies to illustrate the performance of the proposed method. Then in Section 4.5, we produce estimates in a Bayesian manner and include some discussion on the sensitivity of the choice of the prior distribution. We will conclude the results and discuss some future work in Section 4.6. Concluding remarks and more detailed discussion on future research directions are given in Chapter 5.

# Chapter 2

# Augmented Likelihood for Incorporating Auxiliary Information into Left-truncated Data

## 2.1 Introduction

Survival analysis is a branch of statistical methods for failure time or time-to-event data. Examples include analysis of longevity in a population or modeling times to onset or relapse of specific diseases. These kinds of data are often subject to left-truncation. For instance, when considering the duration of survival with a disease, the data are typically collected from participants who had the disease and remained alive at the time of accrual. This implies that their survival time since disease onset has to be longer than the time to study entry, which leads to the so-called left-truncation of the survival times. Failure to consider this sampling condition will lead to invalid inference.

There has been considerable discussion on the estimation of a survival function under left-truncation. The proposed statistical methods can be generally classified as conditional or unconditional approaches (Asgharian et al., 2002). The former develops inference based on the likelihood conditional on the observed truncation times, for example conditional

non-parametric estimation of survival functions for left-truncated data (e.g. Wang et al., 1986; Wang, 1987; Tsai et al., 1987), and conditional partial likelihood methods for the Cox models (e.g. Kalbfleisch and Lawless, 1991; Keiding and Moeschberger, 1992; Wang et al., 1993). Unconditional methods have been discussed by Vardi (1989), Wang (1996), Qin and Shen (2010), and Qin et al. (2011). The unconditional analysis may have an efficiency advantage over a conditional approach when it is possible to parameterize the distribution of the truncation times. For instance Wang (1991) pointed out a non-parametric maximum likelihood estimator (i.e. Vardi, 1989) can be more efficient than the conditional nonparametric product limit estimate when the truncation time follows an uniform distribution.

Sometimes auxiliary information from the subjects that are left out of the study owing to the left-truncation is available, we refer to these subjects as "truncated sample" in the article. Interest lies in combining those with the observed sample for more informative inference. Auxiliary information can be in the form summary-level knowledge regarding the truncated sample, results from previous studies on the same or a similar study population, or census data or statistics from the general population. An example is the Nun Study - Aging Study which contains longevity and cognitive function data for 678 religious sisters in the USA (Snowdon, 2003). The sample is subject to left-truncation because the participants were survivors at the study baseline 1991 from an original population of religious sisters born between 1886 and 1916. An earlier mortality study was conducted on the same population of religious sisters starting in 1965 and continuing until 1989; the details are given in Section 2.3.1. The raw individual-level data for the mortality study are no longer available unfortunately. However the study results are reported in Butler and Snowdon (1996), including the total number of deaths by 1989, the standard mortality ratio of the nun birth-year cohorts versus the US white female population, population size and birth time information of the original set of birth cohorts. It is possible to utilize this auxiliary information to improve efficiency. Another example of such a situation is the Kuakini Honolulu-Asia Aging Study (HAAS), established in 1991 with the objective of investigating dementia and aging processes of a cohort of Japanese-American men (Higuchi et al., 2015). This study was an outgrowth of the Honolulu Heart Program (Huh et al., 2015), which started 25 years earlier in 1965 when a study population of Japanese-American men were recruited. The HAAS participants were survivors of this population and hence their

survival times are subject to left-truncation.

There are two papers (Li and Qin, 1998; Shen, 2014) proposing methods that incorporate information on the size of the "full" study population for estimation of a (conditional) survival function in a nonparametric setting. Li and Qin (1998) proposed a likelihood-based method by adding an extra component to the observed likelihood representing the contribution of the total number of truncated subjects. Shen (2014) proposed a constrained expectation-maximization (EM) algorithm for estimation. In addition to the augmented likelihood approaches, Faucett et al. (2002) considered a method based on multiple imputation using auxiliary variables to deal with the loss in efficiency due to a large right-censored proportion.

Given that most of the literature has discussed methods for incorporating only the total number of truncated individuals, the goal of this chapter is to make use of auxiliary information from the truncated multi-birth-year-cohort sample. Our work is closely related to the idea of (Li and Qin, 1998). However, instead of leaving the distribution of the event time unspecified, we assume a parametric model for it. The auxiliary information on the truncated sample is incorporated by augmenting the conditional likelihood function of observed event times. Great attention is given to the practical issues of what kinds of auxiliary information may be available/useful for a multi-cohort study and how to incorporate these appropriately. The rest of this chapter is organized as follows. In Section 2.2, we describe the notation and the proposed methods that incorporate the aggregated auxiliary information on the number of truncated events. In Section 2.3, we provide the details and results of the application to the Nun Study - Aging Study data, including generating individual level pseudo-data sets based on the auxiliary information and applying the proposed method to the combined data. Section 2.4 gives details of the simulation studies. Concluding remarks and discussion are given in Section 2.5.

## 2.2 Method

### 2.2.1 Left-truncation and Conditional Likelihood

Assume the target population members were born during some calendar intervals, as a combination of $K$ birth cohorts. The birth times for cohort $k$ lie in calendar time intervals $\mathcal{B}_k = [b_k, b_{k+1})$, $k = 1, 2, \ldots, K$, and $\mathcal{B} = \cup_{k=1}^{K} \mathcal{B}_k$. Let $T$ denote the age at death of an individual from the population of interest. Let $B$ and $Y = B + T$ be the calendar times of birth and death respectively.

Suppose a sample of individuals from this population are recruited at a calendar time $a^*$, $a^* > b_{K+1}$, and followed over time for survival up to an administrative censoring time $a^* + \tau^*$. The observed event times are subject to both left-truncation and right-censoring as only those who survived up to $a^*$ could be recruited, and not all of them will reach the death state by $a^* + \tau^*$. Let $\mathcal{S}$ represent the collection of indices for the individuals selected into the sample (i.e. the observed sample), $n$ and $n_k$ be the sample size for the study and for birth cohort $k$ respectively. The observed data of a selected individual $i$ are $\{L_i, X_i, \delta_i\}$, where $L_i = a^* - B_i$ represents the age at accrual time, so $L_i < T_i$, $X_i = \min\{T_i, L_i + \tau^*\}$ is the observed age at the event time and $\delta_i = \mathbb{I}\{X_i = T_i\}$ is the censoring indicator, $i = 1, 2, ..., n$. The conditional likelihood for the sample $\mathcal{S}$ is then

$$\mathcal{L}_0 \;=\; \prod_{i \in \mathcal{S}} P(X_i \mid L_i, T_i > L_i) = \prod_{i \in \mathcal{S}} \frac{f(X_i)^{\delta_i} \mathcal{F}(X_i)^{1-\delta_i}}{\mathcal{F}(L_i)}, \tag{2.1}$$

where $f(\cdot)$and $\mathcal{F}(\cdot)$ are the density and the survival functions of event time $T$ respectively.

### 2.2.2 Auxiliary Information on the Truncated Sample and Augmented Likelihood

Suppose another study was conducted on a much larger sample of individuals that were recruited from the same population during an earlier period $[a, a + \tau)$, where $a + \tau < a^*$.

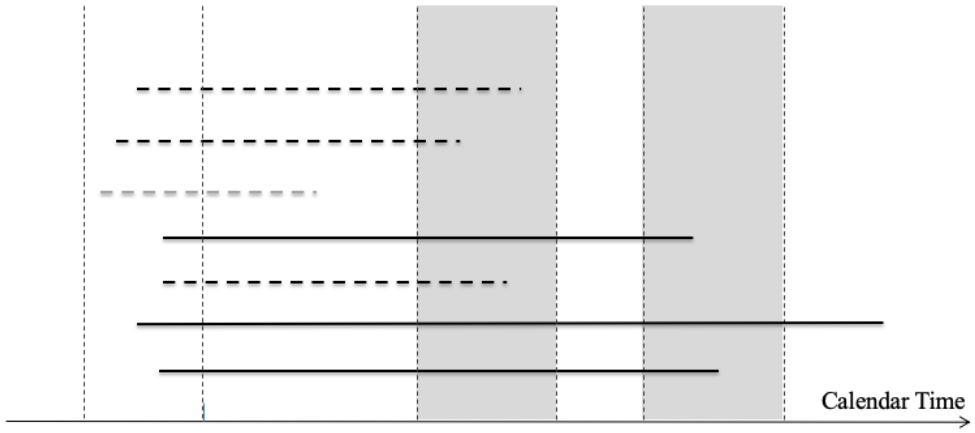


Figure 2.1: An example of the underlying cohorts. The dashed vertical lines represent the birth time window; the shaded region on the left-hand side and right-hand side are the periods for the Nun Study - Mortality Study and the Nun Study - Aging Study respectively. Solid horizontal lines represent individuals who participated in the Aging Study. Dashed horizontal lines represent the left-truncated individuals who died during the study period of the Mortality Study with unknown death times.

Let $N$ indicate the sample size of this study, and $m$ the total number of deaths occurring during this study. The length of this study, $\tau$, can be different for each birth cohort. For instance, in the Nun Study - Mortality Study cohort, the end of follow-up is the minimum of 1989 and the year that participants reached the age of 85, which is identical for participants born in the same calendar year but not so otherwise. The individual-level data for those who had an event during this earlier study period are not obtainable, and let $\mathcal{S}_p$ denote the indices set of this truncated sample. Figure 2.1 gives an example where the shaded region on the left- and right-hand side are the periods for two studies, e.g. the Nun Study - Mortality Study and the Nun Study - Aging Study respectively. The lengths of the segments represent the time to event. Suppose seven individuals were born during a time interval. Six survived to the start of the first study, three died (denoted by dashed lines) during the study period and their death times were not known. The remaining three alive individuals (denoted by solid lines) then entered the second study where lifetime $T$ was observed for two individuals and censored for one individual.

The auxiliary information for the truncated sample $\mathcal{S}_p$ can be of various forms. We focus on the case where the number of deaths for certain age groups are known. However, the idea can be adapted to other types of aggregate information as long as the information can be represented by sample sizes or ranges of the missing observations. We will discuss this further in the Section 2.5. Suppose the $k^{\text{th}}$ birth cohort consists of individuals who were born in the calendar time period $\mathcal{B}_k$. The auxiliary information includes the size of this cohort, $N_k$, and the number of deaths that occurred during $[a, a+\tau)$, denoted by $m_k$, $k = 1, 2, \ldots, K$. Note that $\sum_{k=1}^{K} N_k = N$ and $\sum_{k=1}^{K} m_k = m$. Let $Y_{(i)}^k$ be the $i^{\text{th}}$ order statistic of the calendar death times for individuals who belong to the $k^{\text{th}}$ birth cohort; then knowing $m_k$ is equivalent to knowing $Y_{(m_k)}^k \leq a+\tau, Y_{(m_k+1)}^k > a+\tau$. The augmented likelihood incorporating this sub-group survival information is

$$\begin{aligned}
\mathcal{L} &= \prod_{i \in \mathcal{S}} \frac{f(X_i)^{\delta_i} \mathcal{F}(X_i)^{1-\delta_i}}{\mathcal{F}(L_i)} \prod_{k=1}^{K} P(Y_{(m_k)}^k \leq a+\tau,\ Y_{(m_k+1)}^k > a+\tau) \\
&= \mathcal{L}_0 \prod_{k=1}^{K} \left( \begin{array}{c} N_k \\ m_k \end{array} \right) P\left(Y \leq a+\tau \mid B \in \mathcal{B}_k\right)^{m_k}\ P\left(Y > a+\tau \mid B \in \mathcal{B}_k\right)^{N_k - m_k} \\
&\propto \mathcal{L}_0 \prod_{k=1}^{K} g_k\left(a+\tau\right)^{m_k} \left[1 - g_k\left(a+\tau\right)\right]^{N_k - m_k} \qquad (2.2)
\end{aligned}$$

where

$$g_k\left(u\right) = P\left(Y \leq u \mid B \in \mathcal{B}_k\right) = \int_{b_k}^{b_{k+1}} P\left(T \leq u - b \mid B = b\right) f_B\left(b \mid B \in \mathcal{B}_k\right) db \qquad (2.3)$$

and $f_B\left(b \mid B \in \mathcal{B}_k\right)$ represents the density of the birth time within interval $\mathcal{B}_k$. One may assume that the distribution of the lifetime $T$ is independent of the birth time so that $P\left(T \leq t \mid B = b\right) = 1 - \mathcal{F}(t)$.

It is possible to incorporate more detailed aggregate information, for example, the number of events occurring during specific time intervals. Suppose the study period $[a, a+\tau)$ can be further divided into intervals $a = a_0 < a_1 < \ldots < a_J = a+\tau$, and let $\mathcal{A}_j = [a_{j-1}, a_j)$ where $j = 1, \ldots, J$. Suppose the number of events in the $j$th interval for birth cohort $k$,

$m_{kj}$, is known, and $\sum_{j=1}^{J} m_{kj} = m_k$. Without loss of generality let $\mathcal{A}_{J+1} = [a_j, \infty)$ and $m_{k,J+1} = N_k - m_k$. In this case, the likelihood given in (2.2) becomes

$$\begin{aligned}
\mathcal{L} &= \mathcal{L}_0 \prod_{k=1}^{K} P(Y^k_{(m_{k1})} \leq a_1, Y^k_{(m_{k1}+1)} > a_1, \ldots, Y^k_{(m_{kJ})} \leq a_J, Y^k_{(m_{kJ}+1)} > a_J) \\
&\propto \mathcal{L}_0 \prod_{k=1}^{K} \prod_{j=1}^{J+1} P(Y \in \mathcal{A}_j \mid B \in \mathcal{B}_k)^{m_{kj}} \\
&= \mathcal{L}_0 \prod_{k=1}^{K} \prod_{j=1}^{J+1} \left[ g_k(a_j) - g_k(a_{j-1}) \right]^{m_{kj}} . \qquad (2.4)
\end{aligned}$$

When $J = 1$, the above likelihood reduces to (2.2). Comparing the augmented likelihood (2.2) or (2.4) with the conditional likelihood (2.1), the extra components improve efficiency by incorporating auxiliary information from the truncated sample.

Denote the log-likelihood as

$$l(\theta) = l_0(\theta) + \sum_{k=1}^{K} \sum_{j=1}^{J+1} m_{kj} \log \left[ g_k(a_j) - g_k(a_{j-1}) \right], \qquad (2.5)$$

where a parametric model $f(t;\theta)$ can be assumed for the event time $T$ indexed by parameter vector $\theta$. The maximum likelihood estimator for $\theta$ is obtained by solving the score equations

$$S(\theta) = \frac{\partial l_0(\theta)}{\partial \theta} + \sum_{k=1}^{K} \sum_{j=1}^{J+1} m_{kj} \frac{\partial}{\partial \theta} \log \left[ g_k(a_j) - g_k(a_{j-1}) \right] = 0. \qquad (2.6)$$

Directly maximizing the log-likelihood (2.5) or solving the score function equation (2.6) can be complicated as it involves calculation of the $g_k(\cdot)$ functions. Alternatively, we propose to approximate (2.6) by Monte-Carlo expectation.

### 2.2.3 Monte-Carlo Expectation

Imagine the ideal case where the event times of the truncated sample, $\mathcal{D}_p = \{T_i\}_{i\in\mathcal{S}_p}$, are observable. Let $\mathcal{D} = \{X_i, \delta_i\}_{i\in\mathcal{S}}$ denote the individual-level data for the observed sample. Then the "complete" data are the combination of data from these two samples, $(\mathcal{D}_p, \mathcal{D})$. The complete log-likelihood and the corresponding score functions are

$$l_c(\mathcal{D}_p, \mathcal{D}; \theta) = \sum_{i\in\mathcal{S}} \log\left[f(X_i)^{\delta_i}\mathcal{F}(X_i)^{1-\delta_i}\right] + \sum_{i\in\mathcal{S}_p} \log f(T_i) \tag{2.7}$$

$$S_c(\mathcal{D}_p, \mathcal{D}; \theta) = \sum_{i\in\mathcal{S}} \frac{\partial \log\left[f(X_i)^{\delta_i}\mathcal{F}(X_i)^{1-\delta_i}\right]}{\partial\theta} + \sum_{i\in\mathcal{S}_p} \frac{\partial \log f(T_i)}{\partial\theta}. \tag{2.8}$$

If we take the expectation of the complete score function (2.8) over all of the values for $\mathcal{D}_p$ conditional on the observed data $\mathcal{D}$ and the auxiliary data, these expectations should equal the log-likelihood and the score functions in (2.5) and (2.6). That is,

$$S(\theta) = \mathbb{E}_{\mathcal{D}_p|\mathcal{D}}\left[S_c(\theta) \mid Y^k_{(m_{kj})} \le a_j, Y^k_{(m_{kj}+1)} > a_j, j = 1,\ldots,J, k = 1,\ldots,K\right]. \tag{2.9}$$

Therefore, instead of directly solving (2.6), we solve the conditional expectation of complete score function, as shown in (2.9), equal to 0.

Calculation of the conditional expectation in (2.9) can be complicated, and thus we consider a Monte-Carlo technique. The idea is that for a large enough number $R$, we generate sets of pseudo-values for $\mathcal{D}_p^{(r)}, r = 1, 2, ...R$, satisfying the conditions in (2.9). The average of $S_c(\mathcal{D}_p, \mathcal{D}; \theta)$ of these pseudo-data sets approximates the score function given in (2.6), i.e.,

$$S(\theta) \approx \frac{1}{R}\sum_{r=1}^{R} S_c\left(\mathcal{D}_p^{(r)}, \mathcal{D}; \theta\right), \tag{2.10}$$

given that the assumed model for $T$, with density $f(t;\theta)$, is correct. We know that the event times of participants belonging to the $k^{\text{th}}$ birth cohort satisfy $Y^k_{(m_{kj})} \le a_j, Y^k_{(m_{kj}+1)} > a_j, j = 1,\ldots,J$, and there should be $N_k - n_k$ truncated participants in this birth cohort as

$n_k$ indicates the number of participants who remained alive and made it to the subsequent study. To ensure that the pseudo-data sets meet these constraints, for each $j = 1, \ldots J$ and $k = 1, \ldots K$, we generate $m_{kj}$ calendar birth times using $f\left(b \mid B \in \mathcal{B}_k\right)$, and generate the event times using $f(t \mid T + B \in \mathcal{A}_j, B = b; \widetilde{\theta})$, where $\widetilde{\theta}$ is obtained by maximizing the conditional likelihood (2.1).

### 2.2.4 Efficiency Gain

The proposed procedure for maximizing the augmented log-likelihood (2.5) can be seen as a one-step EM algorithm, where the E-step is the calculation of the conditional expectation of the complete score (2.9) by the Monte-Carlo approximation (2.10), and the M-step is solving $S\left(\theta\right) = 0$. Louis's formula (Louis, 1982) is used to estimate the observed information for the likelihood incorporating the auxiliary information. Specifically we obtain

$$\begin{aligned} I\left(\theta\right) &= \mathbb{E}_{\mathcal{D}_p|\mathcal{D}}\left[-\frac{\partial^2 l_c\left(\mathcal{D}_p, \mathcal{D}; \theta\right)}{\partial\theta\partial\theta^{\mathrm{T}}}\right] \\ &\quad -\mathbb{E}_{\mathcal{D}_p|\mathcal{D}}\left[S_c\left(\mathcal{D}_p, \mathcal{D}; \theta\right) S_c^{\mathrm{T}}\left(\mathcal{D}_p, \mathcal{D}; \theta\right)\right] + S\left(\mathcal{D}; \theta\right) S^{\mathrm{T}}\left(\mathcal{D}; \theta\right). \end{aligned} \tag{2.11}$$

Note that the last term takes values of zero at the maximum likelihood estimates from setting the observed score functions equal to 0.

The conditional expectation is obtained by a Monte-Carlo approximation by simulating pseudo-data sets as before. Then the observed information is approximated as

$$I\left(\theta\right) \approx -\frac{1}{R}\sum_{r=1}^{R}\frac{\partial S_c\left(\mathcal{D}_p, \mathcal{D}; \theta\right)}{\partial\theta^{\mathrm{T}}} - \frac{1}{R}\sum_{r=1}^{R} S_c\left(\mathcal{D}_p, \mathcal{D}; \theta\right) S_c^{\mathrm{T}}\left(\mathcal{D}_p, \mathcal{D}; \theta\right). \tag{2.12}$$

The variance covariance estimate, $\widehat{Var}(\widehat{\theta})$, is calculated by taking the inverse of $I(\widehat{\theta})$. To evaluate the performance of the proposed method, we calculate the estimated variance

reduction for estimating a parameter $\theta_q$ as

$$RV = \frac{\widehat{Var}(\widetilde{\theta}_q) - \widehat{Var}(\widehat{\theta}_q)}{\widehat{Var}(\widetilde{\theta}_q)}, \tag{2.13}$$

where $\widetilde{\theta}_q$ and $\widehat{\theta}_q$ are estimated based on the conditional likelihood (2.1) and the augmented likelihood (2.4) incorporating the auxiliary information respectively.

## 2.3 Application to the Nun Study

### 2.3.1 Sources of the Auxiliary Information

The Nun Study is a prospective longitudinal study started in 1991. The study population consists of all US members of the School Sisters of Notre Dame who were born during the years 1886 to 1916. The participants were a sample of 678 religious sisters from this population and their cognitive function and survival time information were collected during the followup to 2003. We refer to this study as the Nun Study - Aging Study. This sample is subject to left-truncation as the religious sisters needed to be alive at the baseline (year 1991) to be eligible. Another study, the Nun Study - Mortality Study, was conducted on all of the religious sisters who survived to the year 1965 from the same population. The study results reported in Butler and Snowdon (1996) provide auxiliary information on the study population and the truncated sample (deceased individuals). The auxiliary information includes (i) a description of the birth-time distribution of all US members of the School Sisters of Notre Dame born in the period 1886 to 1916; (ii) the size of the study population; there were 2573 religious sisters born during that interval who survived to the year 1965; (iii) the number 1103 of deaths aged 50 to 84 by the end of follow up in 1989; and (iv) the standard mortality ratio (SMR) by age group compared to the US white female population. Figure 2.2 shows the timelines and participants of the Aging Study and the Mortality Study. We think of the 2573 religious sisters participated in the Mortality Study as our “full" underlying cohort, and we view those who passed away by 1991 as the truncated sample and those who remained alive in 1991 and entered the Aging

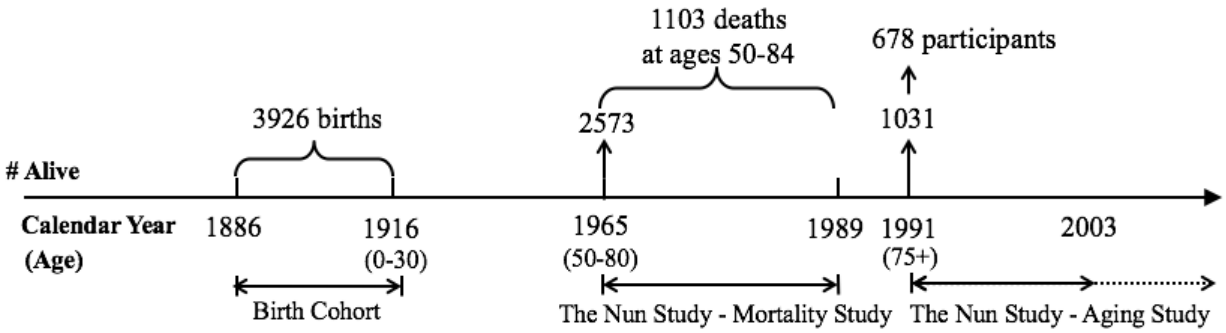


Figure 2.2: The Nun Study timeline.

Study as the observed sample. The auxiliary information is used to estimate the birth time distribution of the population and the number of deaths per year up to 1991 for each birth group. These pseudo mortality data of the truncated sample are then combined with the observed data from the Aging Study to form the "complete" data for the estimation of survival. We describe the details in the following.

### Birth Time Distribution of the Study Population

It was reported that 3926 US members of the School Sisters of Notre Dame were born in $[1886, 1916)$, and that their birth times were approximately uniformly distributed from 1894 to 1910, but not before and after this period (Butler and Snowdon, 1996). We divided this population into $K = 30$ cohorts by birth years $\{b_1, \ldots, b_{30}\} = \{1886, \ldots, 1915\}$, and denote the birth time interval for the $k^{th}$ cohort by $\mathcal{B}_k = [b_k, b_k + 1)$, $k = 1, \ldots, 30$. Thus we assume a piecewise linear trend in numbers of births by year, $\{N_k^*\}_{k=1}^{30}$. To be specific, we assume

$$N_k^* = \phi_0 + \phi_1 (b_k - 1886)_+ - \phi_1 (b_k - 1895)_+ + \phi_1 (b_k - 1909)_+$$

for participants born before 1915, i.e. $k \leq 29$, and set the number of births during 1915 to match the extra small size of the youngest group in the Aging Study. The estimate $\hat{\phi}_1$ is obtained using the average increment in the number of births of the US white female pop-


Figure 2.3: Estimated number of religious sisters who were born in each year from 1886 to 1916 (height of the vertical lines) and number of those who survived to the year 1965 to enter the Nun Study - Mortality Study (positions of the triangles).

ulation from 1910 to 1915, and $\hat{\phi}_0$ is solved by setting $\sum_{k=1}^{30} N_k^* = 3926$. We also assume the probabilities of surviving to 1965 for the religious sisters' population are proportional to those of the white female population in US. Mortality and age distribution information for the US white female population can be obtained from the US Census and the Centers for Disease Control and Prevention websites (e.g. National Vital Statistics Reports 2017; United States National Intercensal Tables: 1900-1990). The ratio of the size of the population aged $1965 - b_k$ in the year 1965 and the total births in the calendar time interval $\mathcal{B}_k$ reflects the chance of survival beyond 1965 for the $k^{th}$ birth cohort, and we denoted it by $\hat{\mathcal{F}}_k$. Given these assumptions, we estimate number of religious sisters who were born in $\mathcal{B}_k$ and survived to the year of 1965 as the following:

$$N_k \quad = \quad \left\lfloor \psi N_k^* \hat{\mathcal{F}}_k + \frac{1}{2} \right\rfloor ,$$

where $\psi$ is a constant that ensures a population size of $\sum_{k=1}^{30} N_k = 2573$, and $\lfloor x \rfloor$ is the largest integer that is smaller than $x$. Figure 2.3 displays the estimated size of each birth cohort. The height of each vertical line represents the number of religious sisters born in the year (i.e., $N_k^*$), and the position of the triangle represents the number of those who survived to 1965 (i.e., $N_k$), which reflects the birth time distribution of the study population. A sensitivity analysis of the underlying assumptions used above is given at the end of Section 2.3.1.

### Estimation of the Number of Truncated Events

The US Vital Statistics Reports contain the current life tables, i.e., the "cross-sectional" life tables, for the US white female population for most of the years from 1965 to 1990 except 1965, 1969, 1979, 1981, and 1990. The tables report for a hypothetical cohort of a certain age in the same year the number of survivors out of 100,000 born alive. Denote the reported number of survivors in the table of calendar year $u$, for age $t$, as $Q(t, u)$. Then the age- and birth-year-specific mortality rate, for a US white female who was born in the year $b_k$, of dying at age $t$ is

$$\begin{aligned} h_0(t, b_k) &= \frac{\#\text{ of deaths at age } t \text{ and calendar year } b_k + t}{\text{exposure-to-risk at age } t \text{ and calendar year } b_k + t} \\ &= \frac{Q(t, b_k + t) - Q(t+1, b_k + t + 1)}{Q(t, b_k + t)}. \end{aligned}$$

For example, for a US white female who was born in the year $b_k = 1910$, the probability of dying at age $t = 60$ is $(Q(60, 1970) - Q(61, 1971))/Q(60, 1970)$, where $Q(60, 1970)$ and $Q(61, 1971)$ can be found in the lifetables of 1970 and 1971 respectively.

Based on the SMR of the participants in the Nun Study - Mortality Study versus the US white female population given in Butler and Snowdon (1996), we estimate the age- and birth-year-specific mortality rate for the population of the religious sisters as

$$h(t, b_k) = h_0(t, b_k) \times \text{SMR}(t, b_k).$$

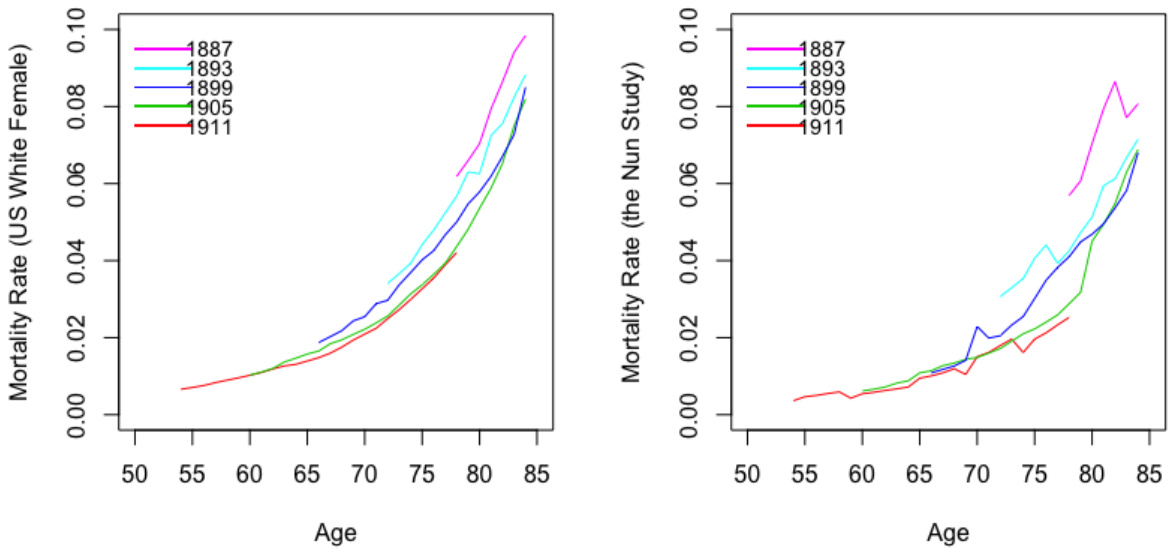


Figure 2.4: Age- and birth-year-specific mortality rates for the US white female population (left) and the participants in the Nun Study - Mortality Study (right) born in the years 1887, 1893, 1899, 1905 and 1911.

Figure 2.4 reports the mortality rates for the US white female population and the estimated rates for the participants in the Nun Study - Mortality Study born in years 1887, 1893, 1899, 1905 and 1911. The number of events occurring in calendar year $a_j$ for participants born in the year $b_k$ can be thus estimated as

$$m_{kj} = \left\lfloor c_k N_k h(a_j - b_k, b_k) \prod_{t=1965-b_k}^{a_j-b_k-1} (1 - h(t, b_k)) + \frac{1}{2} \right\rfloor,$$

where $c_k$ is a constant that ensures $\sum_{j=1}^{J+1} m_{kj} = N_k - n_k$, $b_k = 1886, ..., 1915$, $a_j = 1966, \ldots, \min\{1989, \text{year of age } 84\}$. Table 2.1 gives an example of the number of deaths within each calendar year in a pseudo-data set for the participants who were born in 1900.

Table 2.1: Number of pseudo events by years for participants born in 1900

| Year | Age at Death | $m_{1900,j}$ |
|---|---|---|
| $[1965, 1966)$ | $[65, 66)$ | 2 |
| $[1966, 1967)$ | $[66, 67)$ | 1 |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $[1983, 1984)$ | $[83, 84)$ | 4 |
| $[1984, \infty)$ | $[84, \infty)$ | 22 |

**A Sensitivity Analysis of Underlying Assumptions**

Assumptions were made for the estimation of the numbers of individuals in each age group in the study population, including (i) a piecewise-linear trend in birth for all 3926 religious sisters born in $[1886, 1916)$ in the US; (ii) their chance of survival up to 1965 by year of birth being proportional to that of the US white female population. These assumptions are reasonable to some extent: the auxiliary information on the birth times of all the religious sisters born in the US though incomplete is quite reliable; more than 99% of the religious sisters in the study population were white and the US white female population might be an appropriate reference. It is though important to provide some evidence through sensitivity analyses.

Varying the birth time structure in the initial birth cohort of 3926 religious sisters is a way of changing the estimated age structure of the study population. Increasing the number born later results in higher proportion of younger religious sisters in 1965, consistent with proportionally less mortality from early causes of death such as childbirth. Increasing the number born earlier leads to higher proportion of older religious sisters in 1965, consistent with proportionally less mortality from later life causes of death such as cancers.

Figure 2.5 shows the variations in the estimated age structure of the study population under different scenarios of birth time structure considered for the initial birth cohort. The reference scenario refers to the case when assumptions (i) and (ii) are used for estimation, and the other scenarios correspond to deviations from assumption (i) by inflating the

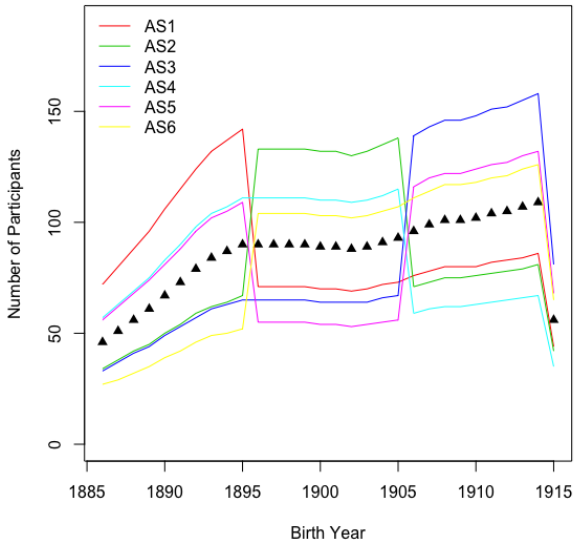


Figure 2.5: Variations of age structure for the study population. Triangles correspond to estimated number of individuals in each age group under assumptions (i) and (ii) indicated in Section 2.3.1, and the other scenarios correspond to deviations from assumption (i) by inflating the number of births for certain period(s).

Table 2.2: Expected number of deaths at ages 50-84 between 1965 and 1989 according to different birth time structures in the initial birth cohort of 3926 religious sisters born in 1886 to 1916. The reference scenario $AS0$ is under assumptions (i) and (ii) given in Section 2.3.1, and the others correspond to deviations from assumption (i) by inflating the number of births for certain period(s).

| Scenario | Period(s) with Inflated Births | | | Expected Deaths |
|---|---|---|---|---|
| AS0 | - | - | - | 1103 |
| AS1 | $[1886, 1895)$ | - | - | 1146 |
| AS2 | - | $[1895, 1905)$ | - | 1171 |
| AS3 | - | - | $[1905, 1915)$ | 998 |
| AS4 | $[1886, 1895)$ | $[1895, 1905)$ | - | 1193 |
| AS5 | $[1886, 1895)$ | - | $[1905, 1915)$ | 1048 |
| AS6 | - | $[1895, 1905)$ | $[1905, 1915)$ | 1072 |

number of births for certain period(s). Table 2.2 reports expected number of deaths at ages 50-84 between 1965 and 1989, calculated under each birth time structure with the use of the SMR and general population mortality rates. They do not vary greatly and are reasonably close to the actual number of deaths (1103 deaths in total) in most cases.

### 2.3.2 Estimating Mortality using the Nun Study Data

First consider the estimation of survival function based on the left-truncated and right-censored data from the Nun Study - Aging Study only. Figure 2.6 compares the estimates of

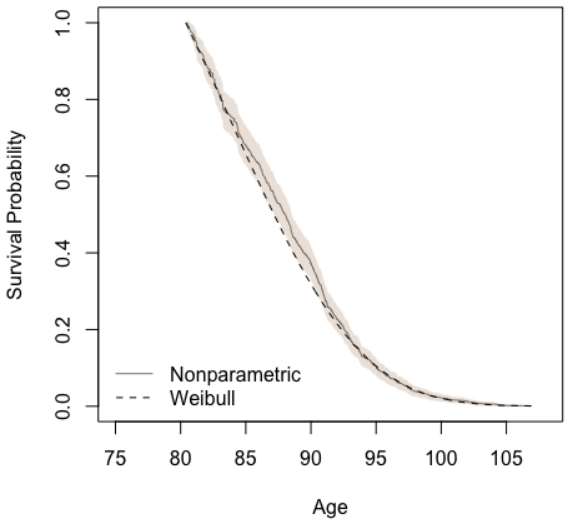


Figure 2.6: Comparison of the estimated conditional survival functions: This plot compares the nonparametric product-limit and the Weibull estimates of the conditional survival function $\mathcal{F}(t \mid T \geq 80)$ based on the observed data from the Nun Study - Aging Study.

the conditional survival function $\mathcal{F}(t \mid T \geq 80)$ based on the non-parametric product limit method and a Weibull distribution. The good agreement between the two curves suggests that a Weibull distribution is suitable for the survival trajectory. Tsai et al. (1987) pointed

out that the occurrence of left-truncation means the problem of small numbers in the risk sets can now affect both the lower and the upper tails of the non-parametric estimate of $\mathcal{F}(t \mid T > T^*)$. They suggest to use a lower bound $T^*$ that is larger than the minimum of the left-truncation times to avoid this difficulty. This is why we estimated $\mathcal{F}(t \mid T > 80)$ instead of $\mathcal{F}(t \mid T > 75)$ using the nonparametric approach.

Next we use the augmented likelihood (2.4) for the estimation of survival function incorporating auxiliary information from the truncated sample. We focus on modeling the survival after age 50; hence the time scale used in the analysis is the actual age minus 50. We consider two parametric distribution models: a Weibull distribution and a two-piece Weibull distribution with a cutoff point at age 75. Because the participants of the Nun Study - Aging Study are aged 75 and older and the estimation of survival before age 75 relies primarily on the truncated sample, the two-piece Weibull model is used to accommodate the potential difference in the mortality patterns before and after that age point. Consider a two-piece Weibull model for the event time $T$ with a single breakpoint $w$,

$$T \sim \begin{cases} \text{Weibull}(\lambda_1, \kappa_1), & \text{if } T \leq w, \\ \text{Weibull}(\lambda_2, \kappa_2), & \text{if } T > w \end{cases}$$

with the density and survival functions being

$$\begin{aligned} f(t) &= \left[\frac{\kappa_1}{\lambda_1}\left(\frac{t}{\lambda_1}\right)^{\kappa_1 - 1} e^{-\left(\frac{t}{\lambda_1}\right)^{\kappa_1}}\right]^{\mathbb{I}\{t \leq w\}} \left[\frac{\kappa_2}{\lambda_2}\left(\frac{t}{\lambda_2}\right)^{\kappa_2 - 1} e^{-\left(\frac{t}{\lambda_2}\right)^{\kappa_2}}\right]^{\mathbb{I}\{t > w\}}, \\ \mathcal{F}(t) &= \left[e^{-\left(\frac{t}{\lambda_1}\right)^{\kappa_1}}\right]^{\mathbb{I}\{t \leq w\}} \left[e^{-\left(\frac{t}{\lambda_2}\right)^{\kappa 2}}\right]^{\mathbb{I}\{t > w\}} \end{aligned}$$

(Jiwani, 2005). To guarantee the connectedness of the survival function, the parameters are subjected to a continuity constraint such that

$$\log(\lambda_1) = \frac{\kappa_2}{\kappa_1}\log(\lambda_2) + \left(1 - \frac{\kappa_2}{\kappa_1}\right)\log(w). \tag{2.14}$$

The proposed augmented likelihood and Monte-Carlo procedure are used for parameter

Table 2.3: Estimates from the Nun Study using Weibull and two-piece Weibull distributions. EST: Maximum likelihood estimates for log parameters; SE: Model-based standard error for MLE; RV: Estimated reduction in variance of the MLE using (2.13).

| | Conditional Likeihood | | Augmented likelihood | | |
|---|---|---|---|---|---|
| | EST | SE | EST | SE | RV % |
| | *Weibull* | | *Weibull* | | |
| $\log \lambda$ | 3.578 | 0.022 | 3.604 | 0.006 | 93.695 |
| $\log \kappa$ | 1.484 | 0.065 | 1.414 | 0.017 | 92.845 |
| | | | *Two-piece Weibull* | | |
| $\log \lambda_1$ | · | · | 3.503 | 0.028 | · |
| $\log \kappa_1$ | · | · | 1.147 | 0.046 | · |
| $\log \lambda_2$ | · | · | 3.609 | 0.005 | 94.189 |
| $\log \kappa_2$ | · | · | 1.463 | 0.019 | 91.622 |

estimation, where $R = 1000$ pseudo-data sets were generated based on the estimated number of death by year for each birth-year-cohort as described in Section 2.3.1. The shape of the estimated survival function from the Weibull distribution is very similar to that from the two-piece Weibull.

Table 2.3 reports the estimated Weibull parameters and their standard errors. The results are based on the conditional likelihood (2.1) of observed data from the Nun Study - Aging Study, and the augmented likelihood (2.4) incorporating the auxiliary information. A Weibull model and a two-piece Weibull model with breakpoint at age 75 are considered. Under the two-piece Weibull model, the estimates of parameters for the $\geq 75$ group are compared with those from the conditional likelihood approach. The variances are estimated by Louis's formula (2.11), and the relative reduction in variance (RV), from augmented likelihood versus the conditional likelihood approaches, is calculated using (2.13). When the yearly information on the truncated sample members are estimated and incorporated in the augmented likelihood, the RV for the estimators of the scale parameters in one- and two-piece Weibull models are 93.7% and 94.2% respectively, and the RV for the shape parameter estimators are 92.8% and 91.6% respectively.

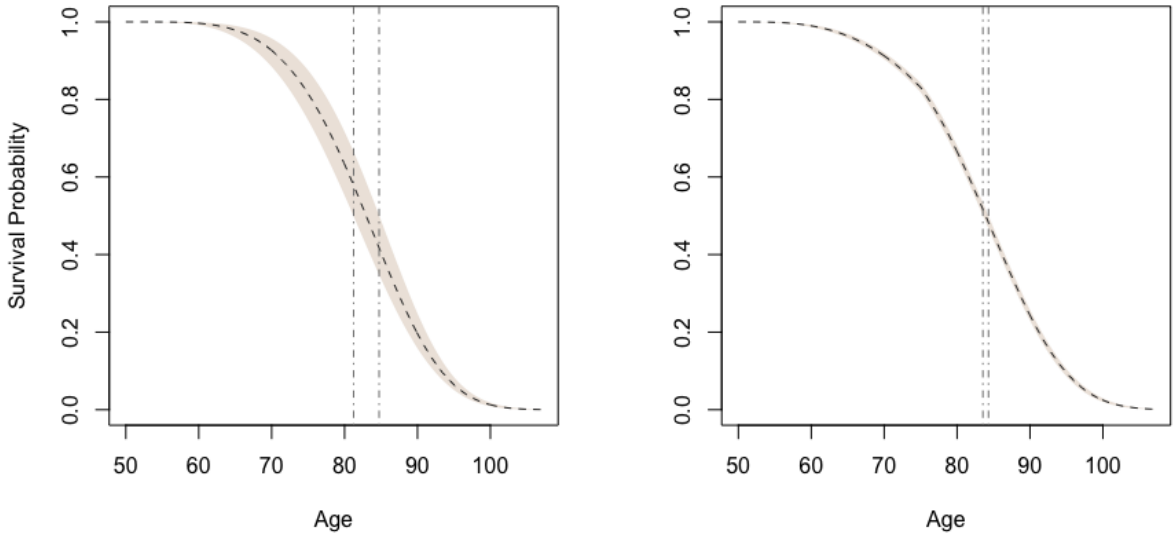


Figure 2.7: Estimates and confidence intervals for the survival functions using a Weibull model based on observed data from the Nun Study - Aging Study and a two-piece Weibull model based on the combined data with auxiliary information in terms yearly number of deaths (right). The vertical lines are the 95% confidence intervals for the median survival age.

Figure 2.7 gives a more straightforward comparison of the estimation accuracy for the survival functions. The left plot is from the conditional likelihood of the observed data and using one-piece Weibull model, while the right plot is from the augmented likelihood and the two-piece Weibull model. The shapes of the estimated survival curves are very similar. The shaded regions represent the 95% confidence intervals for the estimated survival functions. The vertical lines are the 95% confidence intervals for the estimates of the median survival age. The confidence intervals from the augmented likelihood are much narrower than the ones from the conditional likelihood. Estimates for the other quartiles of the survival age can be found in Table 2.4. The differences in the estimated median ages are within one year between the conditional approaches and the proposed methods.

Table 2.4: Estimated quartiles and empirical confidence intervals for quartiles using the yearly auxiliary information for the Nun Study.

| | Conditional, Weibull | | Weibull | | Two-piece Weibull | |
|---|---|---|---|---|---|---|
| | EST | 95% CI | EST | 95% CI | EST | 95% CI |
| Quartile 1 | 77.001 | (74.937, 79.115) | 77.148 | (76.719, 77.575) | 77.675 | (77.236, 78.112) |
| Quartile 2 | 82.957 | (81.243, 84.730) | 83.617 | (83.230, 84.004) | 83.923 | (83.541, 84.307) |
| Quartile 3 | 88.563 | (87.267, 89.951) | 89.786 | (89.358, 90.220) | 89.827 | (89.413, 90.245) |

## 2.4 Simulation Studies

Simulation studies were conducted to evaluate the performance of the method proposed. We considered two sample scenarios that are similar to that of the Nun Study. Sample scenario $\langle 1 \rangle$ is for the simplest case that all individuals were born at time 0 and truncated later at the same time point. For sample scenario $\langle 2 \rangle$, there are three birth cohorts with individuals born during $\mathcal{B}_1 = [0, 0.1)$ , $\mathcal{B}_2 = [0.1, 0.2)$ , and $\mathcal{B}_3 = [0.2, 0.3)$. The birth times follow a uniform distribution within these windows. The left-truncation times are identical for individuals within each birth cohort but differ by 0.1 for those from adjacent birth time intervals. Conditional on the birth times, event times for all of the individuals were generated. At a fixed calendar time, it is assumed that the "observed data" $\mathcal{S}$ were selected by simple random sampling from the survivors at that time. For those who were not included in $\mathcal{S}$, the number of events that occurred within time intervals $\mathcal{A}_j$ was recorded and served as the auxiliary data $\{m_{kj}\}$, whereas the individual-level event times were abandoned. In the estimation procedure, for each pseudo-data set, there were $m_{kj}$ event times generated conditional on the individuals having been born at a uniformly distributed time in interval $\mathcal{B}_k$ and the events occurring within $\mathcal{A}_j$. The pseudo individual-level event times were generated using the MLEs from maximizing the conditional likelihood (2.1).

The results are based on 500 replications and a total sample size of 3000, where the conditional expectations are approximated by averaging over 1000 pseudo-data sets. The distribution used to simulate the event times is the two-piece Weibull with $\lambda_1 = 1, \lambda_2 = 1.09, \kappa_1 = 2, \kappa_2 = 4$, with breakpoint located at the 0.76 quantile of the distribution. For each simulation, we obtain estimates from the conditional likelihood not using any auxiliary

Table 2.5: Empirical bias and reduction in variances using different timelines, types of auxiliary information and left-truncation probabilities. Estimates are obtained using conditional likelihood $\mathcal{L}_0$, augmented likelihood $\mathcal{L}_T$ using total number of deaths and augmented likelihood $\mathcal{L}_I$ using number of deaths in intervals.

| | Sample Scenario ⟨1⟩ | | | | | Sample Scenario ⟨2⟩ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias % | | RV % | | | Bias % | | RV % | | |
| $P(T < L)$ | $\mathcal{L}_T$ | $\mathcal{L}_I$ | $\mathcal{L}_T$ vs $\mathcal{L}_0$ | $\mathcal{L}_I$ vs $\mathcal{L}_0$ | $\mathcal{L}_I$ vs. $\mathcal{L}_T$ | $\mathcal{L}_T$ | $\mathcal{L}_I$ | $\mathcal{L}_T$ vs $\mathcal{L}_0$ | $\mathcal{L}_I$ vs $\mathcal{L}_0$ | $\mathcal{L}_I$ vs. $\mathcal{L}_T$ |
| | | | | *Log Scale*, $\geq 75$, $\log \lambda_2$ | | | | | | |
| 0.2 | 1.415 | 1.434 | 33.797 | 35.893 | 3.051 | 1.128 | 1.167 | 64.517 | 66.689 | 5.685 |
| 0.3 | 1.195 | 1.234 | 62.644 | 64.824 | 5.381 | 0.860 | 0.896 | 84.786 | 86.075 | 7.406 |
| 0.4 | 0.922 | 0.982 | 83.689 | 85.177 | 7.964 | 0.526 | 0.551 | 93.193 | 93.984 | 10.146 |
| 0.5 | 0.632 | 0.693 | 93.514 | 94.263 | 9.563 | 0.352 | 0.392 | 96.396 | 96.933 | 14.295 |
| | | | | *Log Shape*, $< 75$, $\log \kappa_1$ | | | | | | |
| 0.2 | 4.950 | 4.993 | 10.713 | 11.371 | 0.713 | 4.409 | 4.478 | 32.093 | 33.140 | 1.502 |
| 0.3 | 4.351 | 4.437 | 30.263 | 31.551 | 1.795 | 3.781 | 3.850 | 60.172 | 61.310 | 2.795 |
| 0.4 | 3.616 | 3.739 | 58.845 | 59.964 | 2.684 | 2.817 | 2.871 | 80.087 | 80.764 | 3.461 |
| 0.5 | 2.811 | 2.925 | 80.759 | 81.345 | 2.977 | 2.289 | 2.367 | 89.130 | 89.626 | 4.599 |
| | | | | *Log Shape*, $\geq 75$, $\log \kappa_2$ | | | | | | |
| 0.2 | -1.378 | -1.630 | 82.275 | 86.685 | 24.996 | -0.885 | -2.478 | 88.331 | 93.422 | 43.744 |
| 0.3 | -1.053 | -1.670 | 89.280 | 93.515 | 39.726 | 0.280 | -1.758 | 92.683 | 96.894 | 57.401 |
| 0.4 | -0.165 | -1.420 | 93.126 | 96.922 | 55.153 | 2.108 | -1.109 | 93.185 | 98.443 | 73.461 |
| 0.5 | 1.025 | -1.038 | 94.693 | 98.537 | 68.957 | $-^\dagger$ | -0.530 | $-^\dagger$ | 99.057 | $-^\dagger$ |

$^\dagger$When the truncation percentage is high (i.e. $P(T < L) = 0.5$) and the truncation time varies (i.e. sample scenario ⟨2⟩), knowing the total number of deaths alone does not give enough information about the earlier part of the survival distribution to ensure convergence for the estimation of $\log \kappa_2$.

information ($\mathcal{L}_0$) with the ones from the augmented likelihood using the total information ($\mathcal{L}_T$) or interval information ($\mathcal{L}_i$) on the number of truncated individuals.

Table 2.5 reports the average percentage of relative bias and relative reduction in variance estimates over 500 replications. Under the considered settings, the biases in the estimates based on the augmented likelihood are negligible. Meanwhile, the standard errors are substantially smaller than the ones from the conditional likelihood approach. To see the improvement in efficiency, we calculate the relative variance reduction (RV) by comparing the approaches based on the augmented likelihood with the conditional likelihood approach not using any auxiliary information (i.e. $\mathcal{L}_T$ vs $\mathcal{L}_0$ and $\mathcal{L}_I$ vs $\mathcal{L}_0$ ), and then by comparing two augmented likelihood approaches (i.e. $\mathcal{L}_I$ vs $\mathcal{L}_T$). The performance of the

three approaches improves incrementally as more detailed auxiliary information is incorporated. Comparing the results for the two sample scenarios, the proposed method results in a larger efficient gain when truncation time varies (sampling scenario $\langle 2 \rangle$). In addition, the results for different truncation probabilities show that the reduction in variances becomes larger when the truncated samples represent a larger proportion of the whole cohort. With high percentage of truncation (i.e. $P(T < L) = 0.5$) and considerable variation in truncation times (sample scenario $\langle 2 \rangle$), knowing the total number of deaths alone does not in many of the pseudo sets give enough information about the earlier part of the survival distribution to ensure convergence of the estimate of $\log \kappa_2$.

## 2.5 Discussion

In this chapter, we constructed the augmented likelihood incorporating the auxiliary survival information from the truncated sample. From both the simulation and the real data example, incorporating the auxiliary information successfully reduces the loss in efficiency due to the left-truncation. However, the validity of these results rely on the compatibility of the auxiliary information. Careful selection and manipulation of the auxiliary information is very important for the proposed method. For our application on the Nun Study - Aging Study data, the auxiliary survival information is reliable as it comes from an earlier study conducted on the same population. When the auxiliary information is in an alternative form, the way to construct the augmented likelihood becomes different, while its approximation for the value of the likelihood can still be obtained as long as the auxiliary information can be viewed as constraints for the generation of pseudo-data.

To achieve convergence when using Monte-Carlo approximation of the conditional expectations, generating a large enough number of pseudo-data sets is very important. On the other hand, we want to use the smallest number of pseudo-data sets that accomplishes convergence for the sake of time efficiency. In our application, we observed that the value of the mean log likelihood changes while the number of pseudo-data sets increases, and it becomes rather stable when the number of pseudo-data sets is larger than 100. However, the ideal number will depend on the sample size, the underlying distribution, and the

truncation percentage.

In conclusion, incorporation of auxiliary information can substantially increase the efficiency compared to approaches based on observed left-truncated and right-censored data only. The MCEM algorithm makes the calculation of the conditional expectation of the likelihood and its derivatives easier when the closed forms are not obtainable. Future work such as nonparametric approaches, incorporation of risk factors and extension to multistate models are discussed in Chapter 5.

# Chapter 3

# A Mixture Hidden Markov Model with Partially Known Component Memberships

## 3.1 Introduction

The Nun Study is a longitudinal study of aging and cognition among members of a religious congregation, the School Sisters of Notre Dame, living in the United States. Among 1031 eligible religious sisters aged 75 years or older, 678 agreed to participate in the study in 1991-1993, and their cognitive function (e.g., whether the sister has developed dementia or not) was assessed at the baseline and approximately every year thereafter for up to 12 assessments. If a participant passed away during the study, the exact death time was recorded. We are interested in modeling the disease process of dementia. Consider a three-state illness-death model (see Figure 3.1) that contains a disease-free state, a dementia state and a death state. In this process, the death state can be reached from the other two states and the transition from the disease-free state to the dementia state is assumed to be nonreversible. The most widely used multistate model is the Markov model. This kind of model assumes that the transition intensities only depend on the history of the

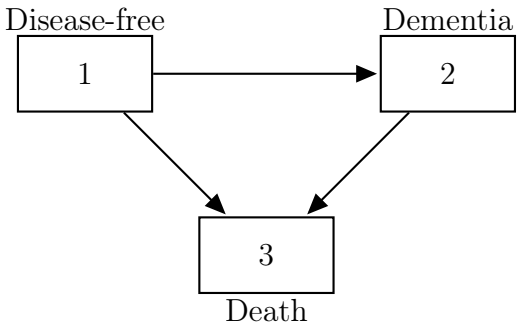


Figure 3.1: A three-state stochastic process for dementia and Alzheimer's disease

process through the current state. The Markov assumption brings great convenience to the derivation of the likelihood function, however it may not be suitable for many applications.

Dementia is a "clinical syndrome" that may result from many diseases, and Alzheimer's disease (AD) is one of the main causes of dementia (Snowdon et al., 1997). In the Nun Study, participants were diagnosed with dementia based on the presence of cognitive impairment (Riley et al., 2002). The criteria include the presence of cognitive impairment in memory and at least one other cognitive domain, functional impairment in activities of daily living, and decline from a previous level of function. In addition, all participants agreed to brain donation after death, thus providing a unique opportunity for the assessment of AD pathology. If an individual has AD, we expect to see plaques and tangles in the brain. These constitute the so-called AD pathology. These pathologic changes will cause problems between or within the neurons and therefore difficulties in cognitive functions. Different criteria have been used to determine whether the development status of the pathology is severe enough to be diagnosed as AD, such as the CERAD neuropathologic criteria and the NIA-RI neuropathologic criteria. The results of the neuropathologic assessment from the Nun Study cohort showed that some of the participants were diagnosed with dementia without development of the AD pathology, which suggests that this group of participants may have suffered from other types of dementia that do not develop AD pathology, for example Parkinson's disease, use of drugs and alcohol, or temporal brain injury, etc. Assuming that AD and other types of dementia have the same incidence may not be reasonable. This motivates us to consider multistate models for mixed types of

disease that give the same clinical expressions. Disease type classification can often be challenging. For example, the neuropathological evaluation of AD obtained from brain biopsy is generally very hard to conduct on a large population. In the Nun Study, the neuropathological assessments, on whether or not AD pathology was present, were conducted on 389 of those deceased. Therefore the disease type is known for some of the participants. The method we propose in this chapter accommodates partially known disease types.

The structure of this chapter is the following. In Section 3.2.1 we introduce the finite mixture hidden Markov model (HMM) and a likelihood-based method for multiple disease types. This method accounts for the heterogeneity in the disease processes and panel observation on disease occurrence. In Section 3.2.2, we extend the mixture HMM to incorporate the pathology data as disease type indicators. In Section 3.3, we apply the mixture HMM to the Nun Study data. In Section 3.4, we conduct simulation studies to ilustrate the performance of the proposed method. Then in Section 3.5, we discuss the identifiability and estimability issues arising in HMMs. In Section 3.6, we use a Bayesian method as an alternative way to obtain the estimates and handle the identifiability issues. Section 3.7 is a discussion.

## 3.2 Method

### 3.2.1 A Finite Mixture Hidden Markov Model

Assume participants were recruited at time $t_0 \geq 0$, and the baseline assessments were taken at accrual. After that, these participants were followed and examined at $K^*$ pre-fixed assessment times $t_1 < t_2 < \ldots < t_{K^*} = \tau$, where $\tau$ is the administrative right censoring time. Suppose the observation process may be terminated due to either the fact that the participant may leave the study at a random time $C$ or death. Let $T_D$ denote the time of death, $T^\dagger = \min\{\tau, C, T_D\}$ and $\delta = \mathbb{I}\{T^\dagger = T_D\}$. Denote the actual number of assessments for a participant as $K = \max\left\{k; t_k \leq T^\dagger; k = 1, \ldots, K^*\right\}$, where $K = K^*$ if the participant was alive and followed until the end of the study, and $K < K^*$ if the participant passed away or dropped out of the study. Suppose the observation disease process $\{Y(t); t \geq 0\}$

is a multistate model with state space $\mathcal{J}_Y = \{1, 2, \ldots, J_Y\}$. The observation of the disease status for a participant, who was alive at $T^\dagger$, is $\{Y(t_0), Y(t_1), \ldots, Y(t_K)\}$, while the set of observed states for the deceased is $\{Y(t_0), Y(t_1), \ldots, Y(t_K), Y(T_D)\}$. Denote $\mathcal{H}_Y(t) = \{Y(s); 0 \leq s \leq t\}$ as the history of the observed state occupancies up to and include time $t$

Let $\{Z(t); t \geq 0\}$ be an underlying stochastic process with state space $\mathcal{J} = \{1, 2, \ldots, J\}$. Denote $\pi = (\pi_1, \pi_2, \ldots, \pi_J)$, where $\pi_j = P(Z(0) = j)$, $j \in \mathcal{J}$. This is the probability for being in an underlying state at the time origin. Let $\mathcal{H}_Z(t) = \{Z(s); 0 \leq s \leq t\}$ be the history of state occupancies of the underlying model up to and include time $t$, the transition intensity from state $i$ to state $j$ at time $t$ is

$$\lambda_{ij}(t \mid \mathcal{H}_Z(t^-)) = \lim_{\Delta t \to 0} \frac{P(Z(t + \Delta t^-) = j | Z(t^-) = i, \mathcal{H}_Z(t^-))}{\Delta t}, \tag{3.1}$$

for $i \neq j$ and $i, j \in \mathcal{J}$. If we assume that $\{Z(t); t \geq 0\}$ is a continuous-time Markov process, then

$$\lambda_{ij}(t \mid \mathcal{H}_Z(t^-)) = \lambda_{ij}(t). \tag{3.2}$$

Let $\Lambda(t)$ be the transition intensity matrix with the $(i, j)$ entries being $\lambda_{ij}(t)$ if $i \neq j$, and the $(i, i)$ entries being $-h_i(t)$, where $h_i(t) = \sum_{j \neq i} \lambda_{ij}(t)$, $i, j \in \mathcal{J}$. For $s < t$, let $\mathrm{P}(s, t)$ be the transition probability matrix where the $(i, j)$ entry is

$$p_{ij}(s, t) = P(Z(t) = j | Z(s) = i), \tag{3.3}$$

for $i, j \in \mathcal{J}$. Under the Markov assumption (3.2), the transition probabilities can be obtained by solving the Kolmogorov forward differential equation

$$\frac{\partial}{\partial t} \mathrm{P}(s, t) \;=\; \mathrm{P}(s, t)\, \Lambda(t). \tag{3.4}$$

If we assume the intensities going from one state to the others are constants over time, i.e. $\Lambda(t) = \Lambda$, the solution to (3.4) is

$$\mathrm{P}(s, t) \;=\; e^{(t-s)\Lambda}. \tag{3.5}$$

This is the so-called time-homogeneous Markov model.

If the underlying model is Markovian, given the current latent state $Z(t)$, the observed state $Y(t)$ does not depend on the history of observation process and the underlying process, that is

$$P\left(Y(t) \mid Z(t), \mathcal{H}_Y\left(t^-\right), \mathcal{H}_Z\left(t^-\right)\right) \quad = \quad P\left(Y(t) \mid Z(t)\right). \tag{3.6}$$

In hidden Markov models, the observed state occupation at a time point is decided by the underlying state that the participant is in at the same time point. The connection between the observation model and the underlying model is represented by the emission probability matrix $\mathrm{E}_{J\times J_Y}$, where the $(j,u)$ entry is $e_{ju}(t) = P\left(Y(t) = u | Z(t) = j\right), j \in \mathcal{J}, u \in \mathcal{J}_Y$. If we assume that the relationship between the observation and underlying models is invariant in time, and that each underlying state is associated with one of the states of the observation model, the emission probabilities become emission indicators, where the $(i,j)$ entry becomes

$$e_{ij} = \begin{cases} 1, & \text{if state } i \text{ of } Z(t) \text{ is an underlying state of state } j \text{ of } Y(t) \\ 0, & \text{otherwise.} \end{cases}$$

Note that a non-hidden Markov model can be viewed as a special Hidden Markov Model where the emission probability matrix is a $J_Y$-dimensional identity matrix.

If all of the participants were alive at the end of the study or censored before deaths occurred, and all of the participants were recruited at the time origin of the process, i.e. $t_0 = 0$, the likelihood for the HMM with panel observation could be written in a matrix-multiplication form (Titman and Sharples, 2010):

$$\mathcal{L}\left(\Theta | Y\right) \quad = \quad \phi_0 \mathrm{Q}_1 \mathrm{Q}_2 \ldots \mathrm{Q}_K \mathbf{1}, \tag{3.7}$$

where $\mathrm{Q}_k, k = 1, 2, \ldots, K$ are $J \times J$ matrices with the $(i,j)$ entry being

$$e_{jy_k} p_{ij}(t_{k-1}, t_k),$$

and $\phi_0 = (\pi_1 e_{1y_0}, \pi_2 e_{2y_0}, \ldots, \pi_J e_{Jy_0})$ is a row vector of entries $\phi_{0,j} = P(Z(0) = j, Y(0) = y_0) = P(Y(0) = y_0 \mid Z(0) = j)P(Z(0) = j)$, $j \in \mathcal{J}$. and $\mathbf{1} = \left[ \begin{array}{cccc} 1, & \ldots, & 1 \end{array} \right]^{\mathrm{T}}_{1\times J}$. The term $p_{ij}(t_{k-1}, t_k)$ is the probability of being in certain state $j$ at assessment time $t_k$ conditional on being in state $i$ at assessment time $t_{k-1}$. Multiplying by $e_{jy_k}$, only transitions that can lead to the observed $Y(t_{k-1})$ and $Y(t_k)$ are kept. It can be easily seen that

$$\mathrm{Q}_k \;=\; \mathrm{P}\left(t_{k-1}, t_k\right) \left[ \begin{array}{cccc} e_{1y_k} & 0 & \cdots & 0 \\ 0 & e_{2y_k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_{Jy_k} \end{array} \right]$$

Modifications are required when the chances of recruitment depend on the state occupancy at the beginning of the study. Consider the case where participants are randomly selected from the individuals who are not in an absorbing state, e.g., death, at the time of recruitment. Define the probability for being in state $j$ of the underlying model at time $t_0$ given the individual is alive as:

$$p_j\left(t_0\right) \;=\; P\left(Z\left(t_0\right) = j \mid \text{alive at } t_0\right),$$

then

$$p_j\left(t_0\right) = \frac{\sum_k p_{kj}(0, t_0)\pi_k}{g\left(\pi, t_0\right)}, \tag{3.8}$$

where

$$g\left(\pi, t_0\right) \;=\; \left[ \begin{array}{cccc} \pi_1, & \pi_2, & \ldots, & \pi_J \end{array} \right] \mathrm{P}\left(0, t_0\right) \left[ \begin{array}{c} I_1 \\ I_2 \\ \vdots \\ I_J \end{array} \right], \tag{3.9}$$

and $I_j = \mathbb{I}\{\text{state } j \text{ is not absorbing}\}$.

Given the sampling condition, let $\phi(t_0, y_0)$ be a row vector of the joint probabilities of

being in observed state $y_0$ and potential underlying state $j$ at the time of recruitment $t_0$, $j \in \mathcal{J}$. Its $j^{\text{th}}$ entry is

$$P(Z(t_0) = j, Y(t_0) = y_0 \mid \text{alive at } t_0) \;=\; e_{jy_0} p_j(t_0).$$

Then

$$\begin{aligned}
\phi(t_0, y_0) &= \left[\, p_1(t_0), \;\; p_2(t_0), \;\; \ldots, \;\; p_J(t_0) \,\right] \begin{bmatrix} e_{1y_0} & 0 & \cdots & 0 \\ 0 & e_{2y_0} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_{Jy_0} \end{bmatrix} \\
&= g(\pi, t_0)^{-1} \pi \mathrm{Q}_0, \qquad (3.10)
\end{aligned}$$

where $\mathrm{Q}_0$ is a $J \times J$ matrix with the $(i, j)$ entry being $e_{jy_0} p_{ij}(0, t_0)$. Then the likelihood for panel observations with non-identical recruitment time $t_0$ is

$$\mathcal{L} \;=\; \phi(t_0, y_0)\mathrm{Q}_1 \ldots \mathrm{Q}_K \mathbf{1} = g(\pi, t_0)^{-1} \pi \mathrm{Q}_0 \mathrm{Q}_1 \ldots \mathrm{Q}_K \mathbf{1}. \qquad (3.11)$$

If death time is observed for a participant, the contribution for the death observation will be the probability of being in any of the non-absorbing states prior to death, and going to the absorbing state right at $T_D$. So we consider a $J \times J$ matrix $\mathrm{Q}_D$ with the $(i, j)$ entry being

$$e_{jy_{t_D}} \left\{ \sum_{l=1}^{J} p_{il}(t_K, t_D) \lambda_{lj}(t_D) \right\}.$$

The likelihood for panel observations with non-identical recruitment time and exactly observed death can thus be written as

$$\mathcal{L} \;=\; g(\pi, t_0)^{-1} \pi \mathrm{Q}_0 \mathrm{Q}_1 \ldots \mathrm{Q}_K \left\{ \mathrm{Q}_D \right\}^{\delta} \mathbf{1}, \qquad (3.12)$$

It can be shown that

$$\mathrm{Q}_D \;=\; \mathrm{P}\left(t_K, t_D\right)\Lambda \begin{bmatrix} e_{1y_{t_D}} & 0 & \cdots & 0 \\ 0 & e_{2y_{t_D}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e_{Jy_{t_D}} \end{bmatrix}.$$

Consider the case that the sampled participants are at risk for several similar diseases, characterized by stochastic processes $Z^{(1)}(t), Z^{(2)}(t), \ldots, Z^{(M)}(t)$ correspondingly. Let $\mathcal{J}^{(m)}$ be the state space of $Z^{(m)}(t), m = 1, 2, \ldots, M$. Suppose these diseases are not differentiable by the stages of the clinical symptoms, i.e. they share the same observed disease status $Y(t)$. Then the underlying model takes the form of one of the underlying processes, that is

$$Z(t) \;=\; \sum_{m=1}^{M} \mathbb{I}\{\text{Type } m\} Z^{(m)}(t), \tag{3.13}$$

where $Z^{(m)}(t), m \in \{1, 2, \ldots, M\}$ are distinct Markov models. The state space of $Z(t)$ will be $\mathcal{J}^{(m)}$ if the underlying disease process is of type $m$. Let $\pi_j^{(m)}$ and $\lambda_{ij}^{(m)}(t)$ be the initial probability and the transition intensities for $Z^{(m)}(t)$, and use superscripts $(m)$ for quantities that are calculated using $\pi_j^{(m)}$ and $\lambda_{ij}^{(m)}(t)$. The likelihood for finite mixture hidden Markov model is

$$\mathcal{L} \;=\; \sum_{m=1}^{M} \psi_m g\left(\pi^{(m)}, t_0\right)^{-1} \pi^{(m)} \mathrm{Q}_0^{(m)} \mathrm{Q}_1^{(m)} \ldots \mathrm{Q}_K^{(m)} \left\{\mathrm{Q}_D^{(m)}\right\}^{\delta} \mathbf{1}, \tag{3.14}$$

where $\psi_m$ is the probability of getting disease type $m$, i.e. the mixture probability for $Z^{(m)}(t)$.

Table 3.1: Numbers of dementia and/or death cases in the Nun Study sample

| | Died | Censored | Total |
|---|---|---|---|
| Dementia | 285 | 12 | 297 |
| No dementia | 321 | 60 | 381 |
| Total | 606 | 72 | 678 |

Table 3.2: Autopsy results for the pathological evaluation of AD for participants in the Nun Study

| | CERAD neuropathologic criteria | | | | |
|---|---|---|---|---|---|
| | Definite | Probable | Possible | No | Missing |
| Dementia | 92 | 65 | 19 | 23 | 86 |
| No dementia | 35 | 71 | 20 | 64 | 231 |
| Total | 263 | | 126 | | 217 |

### 3.2.2 Partially Known Disease Type Information

The Nun Study data involves longitudinal follow-up data for 678 individuals up to 12 assessments, where the assessments are approximately annual. Clinical assessments of dementia were recorded yearly for each individual up to their death or drop-out from the study. The average follow-up time is 8.59 years. Dementia was identified based on the presence of cognitive impairment (Riley et al., 2002). Numbers of dementia and/or death cases in the Nun Study sample are given in Table 3.1. We use the CERAD neuropathologic criteria as the indicator of the AD pathology. Based on counts of plaques, the certainty of AD is divided into four levels: 1 - definite, 2 - probable, 3 - possible and 4 - no; see Table 3.2. Here we consider the levels "definite" and "probable" as the indicator of AD.

Given the pathology data, differentiating the underlying models for dementia with AD pathology versus dementia without AD pathology is likely to help with understanding the disease process. Consider a HMM with underlying model being a mixture of two types of disease; see Figure 3.2, where $\{Y(t)\}$ is the observed disease process, $\{Z^{(1)}(t)\}$ is the

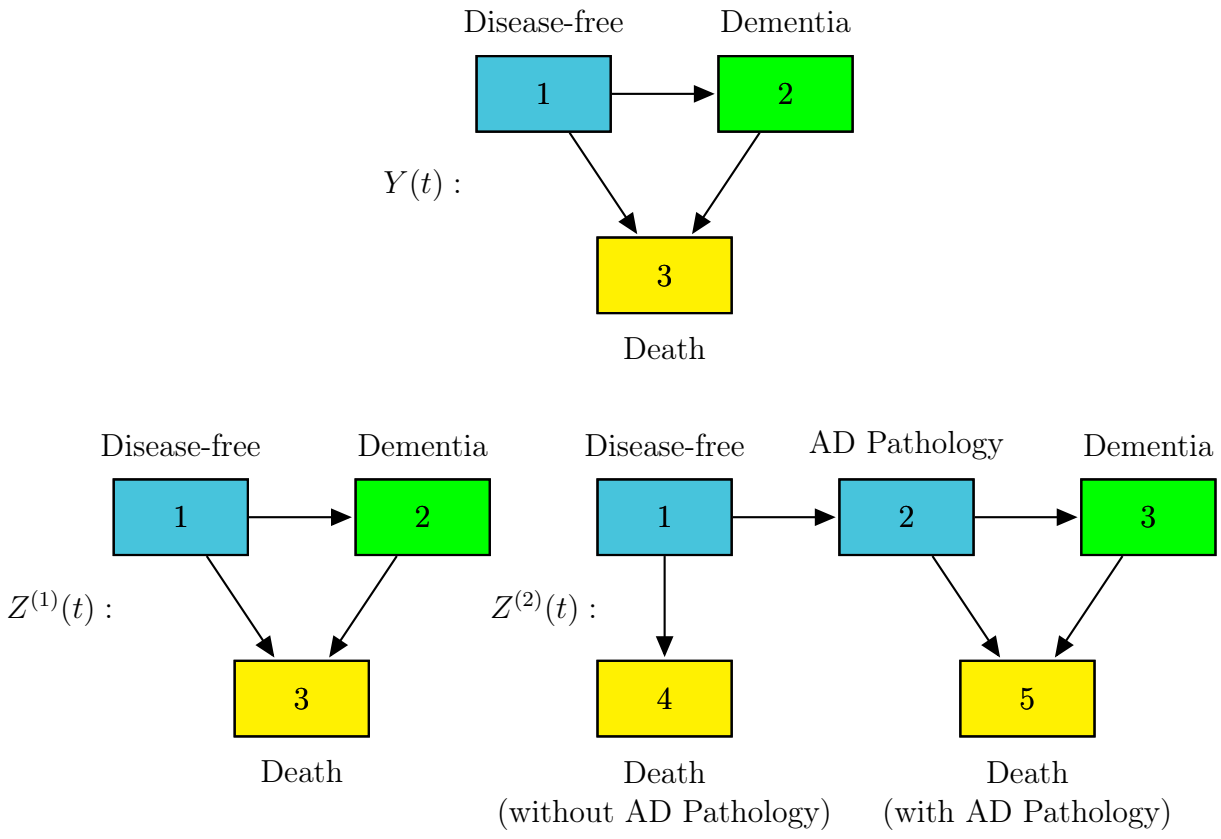


Figure 3.2: A mixture hidden Markov model (HMM) for dementia and Alzheimer's disease. The observation process $Y(t)$ is an illness-death process; the underlying process is a mixture of two types of dementia: $Z^{(1)}(t)$ without and $Z^{(2)}(t)$ with coexistence of AD pathology.

process for dementia type 1 that is not caused by AD pathology, while $\{Z^{(2)}(t)\}$ is for dementia type 2 with coexistence of AD pathology. We assume all participants will follow one of these two disease processes. For participants with AD-pathology, we add a new state representing the development of the AD pathology and assume the dementia state will be reached after the pathology is present. The color in Figure 3.2 represent the correspondence of the state(s) in the underlying model to the states in the observation model. For $Z^{(1)}(t)$, this underlying disease process has a one-to-one mapping to the observation process. For $Z^{(2)}(t)$, the disease process is assumed to be a 5-state Markov model where the healthy

state and the pathology state can not be differentiated from the observed data, and the death state of $Y(t)$ is related to both the disease-free death and the death occurred after development of pathology. Supposing the observed cognition data are based on a mixture of the two types of underlying disease processes, our goal is to estimate the prevalence for each state of the observation disease processes over time.

The onset of the disease pathology cannot be monitored continuously over time, and only current status information can be obtained upon the death with brain autopsy. Consider the Nun Study case where the underlying model $Z(t)$ is a mixture of two Markov models. Let $\mathrm{E}^{(1)}$ and $\mathrm{E}^{(2)}$ denote the emission probability matrices of observation process $Y(t)$ given each of the latent disease processes $Z^{(1)}(t)$ and $Z^{(2)}(t)$ respectively. Based on the structures for the two underlying disease types, the emission probability matrices for the two disease types are

$$\begin{aligned} \mathrm{E}^{(1)} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \\ \mathrm{E}^{(2)} &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Let $X$ be the autopsy result, and $\Delta = \mathbb{I}\{X \text{ is not missing}\}$ be the indicator of observing $X$. For individuals whose pathology data are missing, their component membership is unknown and the likelihood in (3.14) can be directly used:

$$\mathcal{L}_0 = \sum_{m=1}^{2} \psi_m g\left(\pi^{(m)}, t_0\right)^{-1} \pi^{(m)} \mathrm{Q}_0^{(m)} \mathrm{Q}_1^{(m)} \ldots \mathrm{Q}_K^{(m)} \left\{\mathrm{Q}_D^{(m)}\right\}^{\delta} \mathbf{1}. \tag{3.15}$$

where $\mathrm{Q}_k^{(m)}$ is the contribution for the $k^{\text{th}}$ panel observation with emission probability

matrix $\mathrm{E}^{(m)}$.

If an autopsy was conducted and gave positive results, the participant must belong to disease type 2 and have gone through the pathology state (state 2). The likelihood contribution from such a participant is

$$\mathcal{L}_{11} \;=\; \psi_2 g\left(\pi^{(2)}, t_0\right)^{-1} \pi^{(2)} \mathrm{Q}_0^{(2)} \mathrm{Q}_1^{(2)} \ldots \mathrm{Q}_K^{(2)} \left\{\mathrm{Q}_D^{(2)}\right\}^{\delta} r, \tag{3.16}$$

where the right-multiplied vector $r$ controls the number of possibilities to be summed up, and we want to guarantee that state 2 is experienced at some time point. Since our model is a strictly progressive model, define

$$r_i \;=\; \mathbb{I}\{\text{state } i \text{ can be reached from state 2}\}\ ,$$

which gives

$$r \;=\; \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

Multiplying by this vector will eliminate all of the paths that end in states that do not go through state 2.

If an autopsy was conducted and gave negative results, there are two possibilities: 1 - the participant belongs to disease type 1; 2 - the participant belongs to disease type 2 but died before the onset of the AD pathology. The likelihood for such participants is a mixture of contributions from $Z^{(1)}(t)$ process and $Z^{(2)}(t)$ process with a state 1 to state 4 path, i.e.

$$\mathcal{L}_{10} \;=\; \sum_{m=1}^{2} \psi_m g\left(\pi^{(m)}, t_0\right)^{-1} \pi^{(m)} \mathrm{Q}_0^{(m)} \mathrm{Q}_1^{(m)} \ldots \mathrm{Q}_K^{(m)} \left\{\mathrm{Q}_D^{(m)}\right\}^{\delta} c_m. \tag{3.17}$$

where $c_1 = \mathbf{1}$ and $c_2 = \mathbf{1} - r$.

Finally the likelihood function incorporating the partially known component membership based on available pathology data is a combination of above three cases. For an individual with definite or probable pathology, the likelihood contribution is (3.16). For individuals with possible or no pathology, use (3.17). For censored subjects or subjects who did not have the pathology information, the likelihood contribution is (3.15). That is

$$\mathcal{L} = \mathcal{L}_0 \mathbb{I}\{\Delta = 0\} + \mathcal{L}_{10} \mathbb{I}\{\Delta = 1, X = 0\} + \mathcal{L}_{11} \mathbb{I}\{\Delta = 1, X = 1\}. \tag{3.18}$$

The vector of parameters is $\Theta = \left(\psi_m, \pi^{(m)}, \lambda^{(m)}; m \in \{1, 2\}\right)$. The maximum likelihood estimates for the parameters are obtained by maximizing the likelihood function (3.18) while subject to

$$\begin{aligned} \left( \widehat{\psi}_1, \ \widehat{\pi}^{(1)}, \ \widehat{\pi}^{(2)}, \ \widehat{\lambda}^{(1)}, \ \widehat{\lambda}^{(2)} \right) &= \text{argmax } \mathcal{L}\left(\Theta | Y\right) \\ \text{s.t. for } m \in \{1, 2\}, \ \lambda_{i,j}^{(m)} &\geq 0, \text{ for } i, j \in \mathcal{J}^{(m)} \\ \psi_1 &\in [0, 1] \\ \pi_j^{(m)} &\in [0, 1] \\ \sum_{j \in \mathcal{J}^{(m)}} \pi_j^{(m)} &= 1. \end{aligned} \tag{3.19}$$

## 3.3 Results from the Nun Study

Table 3.3 gives the maximum likelihood estimates of the parameters using formula (3.18). The probability for being in the pathology state at the initial age 75 was fixed at zero. This is to deal with an identifiability issue arising in HMM. We will discuss this in Section 3.5. The 95% confidence intervals for the log intensities are calculated using the Wald's method; and the model-based standard errors are the square roots of the diagonal elements in the inverse of the negative information matrix. The bounds for the 95% CI for the intensities are obtained by taking exponential of the bounds for the log intensities. Being in the dementia state after the development of the AD pathology at time 0 is allowed.

Table 3.3: Estimated transition intensities for dementia from the Nun Study data using mixture HMM. Parameters are estimated while fixing $\pi_2^{(2)} = 0$. Confidence intervals (CI) are obtained by taking exponentials of the CI bounds for the log transitition intensities, which are calculated based on the MLEs and model-based standard errors for the log transitition intensities.

| | | EST | 95% CI |
|---|---|---|---|
| *Disease Type 1* | | | |
| Disease-free to Dementia | $\lambda_{12}^{(1)}$ | 0.055 | (0.024, 0.130) |
| Disease-free to Death | $\lambda_{13}^{(1)}$ | 0.088 | (0.050, 0.154) |
| Dementia to Death | $\lambda_{23}^{(1)}$ | 0.377 | (0.290, 0.489) |
| *Disease Type 2* | | | |
| Disease-free to Pathology | $\lambda_{12}^{(2)}$ | 0.153 | (0.075, 0.311) |
| Disease-free to Death | $\lambda_{14}^{(2)}$ | 0.033 | (0.004, 0.288) |
| Pathology to Dementia | $\lambda_{23}^{(2)}$ | 0.122 | (0.094, 0.158) |
| Pathology to Death | $\lambda_{25}^{(2)}$ | 0.088 | (0.066, 0.117) |
| Dementia to Death | $\lambda_{35}^{(2)}$ | 0.254 | (0.223, 0.289) |
| *Other Parameters* | | | |
| Mixture probability | $\psi_1$ | 0.262 | (0.123, 0.474) |
| Initial probability of $Z^{(1)}(t)$ | $\pi_1^{(1)}$ | 0.977 | (0.028, 1.000) |
| Initial probability of $Z^{(2)}(t)$ | $\pi_1^{(2)}$ | 0.974 | (0.825, 0.997) |

According to the result, around 26% of the participants were classified as type 1, which relates to other types of dementia than AD, with 95% confidence interval (12.3%, 47.4%). The intensity of going to the dementia state with presenting of AD pathology is more than twice as high comparing to participants without the AD pathology, i.e. 0.122 vs. 0.055. It takes an average of 6.536 years for type 2 participants to develop the AD pathology and then 8.197 years to proceed to dementia, comparing to 18.182 years for type 1 participants to go from healthy directly to dementia. For both types of participants, being diagnosed with dementia increases their transition intensities to the death state. The mortality rate is 4.284 times higher for type 1 participants and 7.697 times higher for type 2 participants.

Figure 3.3 gives the estimated prevalence of dementia for participants who were at the

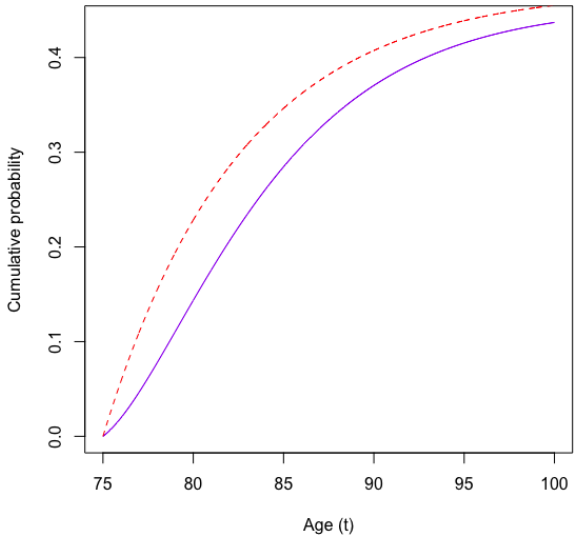


Figure 3.3: Estimated cumulative probability for dementia for religious sisters who were in healthy state at age 75. The red dashed curve is from a time-homogeneous illness-death model; the purple solid curve is from the proposed mixture HMM.

disease-free state at age 75. The x-axis is age. The y-axis is the cumulative probability for being clinically diagnosed with dementia, and the calculation of this probability is given in Appendix A.2. The purple solid curve is the result from the proposed mixture HMM, and the red dashed curve is from time-homogeneous illness-death model. When using a single time-homogeneous illness-death model, the shape of the prevalence curve has to be from an exponential distribution, while our proposed method allows a more flexible slope that can go steeper or flatter at different ages. Our method gives an overall smaller prevalence of dementia than the homogeneous model.

Figure 3.4 compares prevalence of dementia between the two disease types. The left-hand-side curve is for non-AD dementia, and the right-hand-side curve is for dementia with AD pathology. According to these curves, dementia due to AD has a lower incidence rate at younger ages (75 - 85), while this rate is increasing as age becomes older and exceeds that of non-AD dementia eventually. The prevalence for dementia are around 40% and

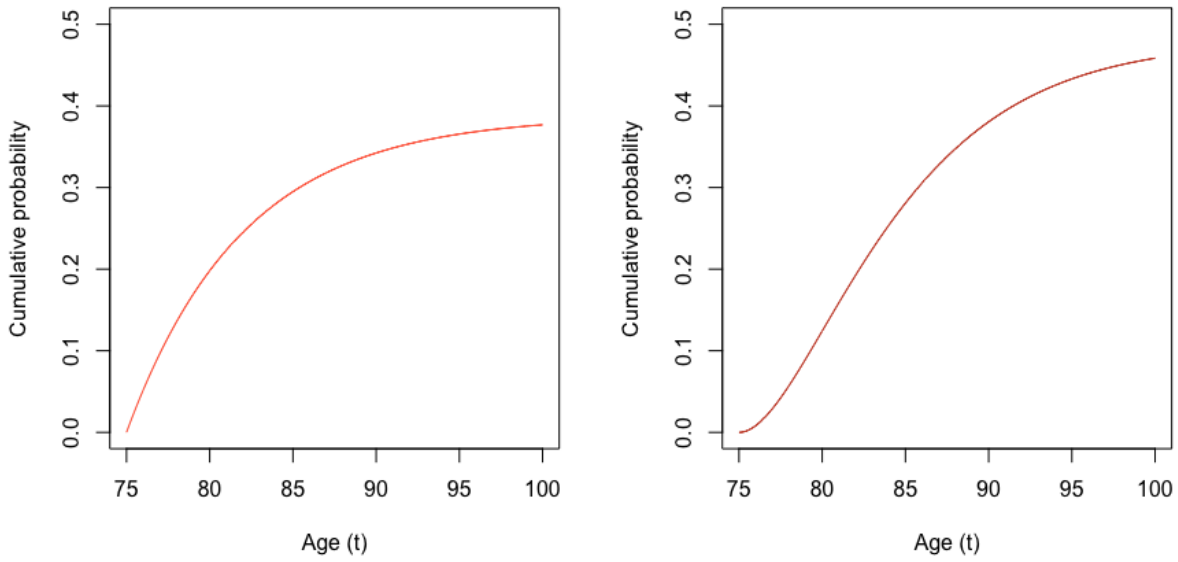


Figure 3.4: Cumulative probability from the Nun Study for two types of dementia type 1 (left) without and type 2 (right) with dementia with AD pathology for people who are healthy at age 75

45% for non-AD and AD groups respectively.

## 3.4 Simulation Studies

To evaluate the performance of the proposed mixture HMM, we conducted simulation studies. Consider the same observation and underlying models as shown in Figure 3.2, and assume that the "study period" is taken from time 0 to time 1, and a sequence of observations of $Y(t)$ are recorded at $\{0, 0.25, 0.5, 0.75, 1\}$. Let the chance of having either disease type be 0.5 and the initial probabilities for the two disease types be $\pi^{(1)} = (0.7, 0.3, 0)$ and $\pi^{(2)} = (0.4, 0.3, 0.3, 0, 0)$ respectively. The true values of the transition intensities are calculated by setting the overall right censoring rate to 0.2 while assuming $\lambda_{12}^{(1)}/\lambda_{13}^{(1)} = 2$,

$\lambda_{23}^{(1)}/\lambda_{13}^{(1)} = 1.5$ for disease type 1, and $\lambda_{12}^{(2)}/\lambda_{14}^{(2)} = 2.2$, $\lambda_{25}^{(2)}/\lambda_{14}^{(2)} = 1.8$, $\lambda_{35}^{(2)}/\lambda_{14}^{(2)} = 2.5$ for disease type 2. The recruitment time is 0 for every participant. The values of the transition intensities of the two types of disease processes are given below:

$$\Lambda_1 = \begin{bmatrix} -4.302 & 2.383 & 1.191 \\ 0 & -1.787 & 1.787 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\Lambda_2 = \begin{bmatrix} -2.621 & 1.802 & 0 & 0.819 & 0 \\ 0 & 0 & -3.931 & 2.457 & 1.474 \\ 0 & 0 & 0 & -2.047 & 2.047 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For each individual, we first generate the mixture component membership using a Bernoulli distribution with mean 0.5, then generate panel observations for this individual using underlying model $Z^{(m)}(t), m \in \{1, 2\}$. If this individual is censored at the end of follow-up, we discard the membership information and treat it as unknown, otherwise, we have 80% chance for keeping the membership as known.

Table 3.4 gives simulation results based on 700 replications and sample size 500. In our simulation scenario, the identifiability and estimability issues arise when the proposed method is applied. These are common problems in hidden Markov models (Titman and Sharples, 2010). Therefore the maximization procedure was conducted under the constraint that the ratio $\pi_2^{(2)}/\pi_1^{(2)}$ is known. More discussion regarding the identifiability and estimability issues for the HMM will be given in Section 3.5. The standard errors in the table are the empirical standard errors of the estimates from 700 replications. For each replication, the 95% confidence intervals are calculated using the estimates and model-based standard errors from that replication, and the portions of confidence intervals that contain the true value are reported in Table 3.4. According to the simulation results, the average values of the estimates are close to the true values of the parameters. The coverage probabilities of the 95% confidence intervals are close to 0.95, which suggests that the estimated standard deviations are reasonable.

Table 3.4: Simulation results based on 700 replications with sample size 500. The MLEs are obtained while assuming known $\pi_2^{(2)}/\pi_1^{(2)}$. EST: average estimates; $\text{SE}_{\text{emp}}$: empirical standard errors; CI.cover: coverage probability for 95% confidence intervals.

| | TRUE | EST | $\text{SE}_{\text{emp}}$ | CI.cover |
|---|---|---|---|---|
| *Disease Type 1* | | | | |
| $\lambda_{12}^{(1)}$ | 2.383 | 2.984 | 0.669 | 97.348 |
| $\lambda_{13}^{(1)}$ | 1.191 | 0.754 | 0.514 | 99.242 |
| $\lambda_{23}^{(1)}$ | 1.787 | 1.838 | 0.299 | 92.614 |
| *Disease Type 2* | | | | |
| $\lambda_{12}^{(2)}$ | 1.802 | 2.202 | 0.728 | 99.432 |
| $\lambda_{14}^{(2)}$ | 0.819 | 0.745 | 0.645 | 94.886 |
| $\lambda_{23}^{(2)}$ | 2.457 | 2.251 | 0.515 | 94.129 |
| $\lambda_{25}^{(2)}$ | 1.474 | 2.135 | 0.453 | 87.500 |
| $\lambda_{35}^{(2)}$ | 2.047 | 2.081 | 0.364 | 91.667 |
| *Other Parameters* | | | | |
| $\psi_1$ | 0.500 | 0.443 | 0.048 | 94.886 |
| $\pi_1^{(1)}$ | 0.700 | 0.661 | 0.045 | 98.295 |
| $\pi_1^{(2)}$ | 0.400 | 0.416 | 0.020 | 93.561 |

## 3.5 Parameter Identifiability and Estimability in Mixture Hidden Markov Models

Two common problems in hidden Markov models (HMM) are the identifiability and the estimability issues. As pointed out by Titman and Sharples (2010), these issues are more likely due to the incomplete nature of the observation scheme and the potential complexity of the models. To show the identifiability for the proposed mixture HMM is highly nontrivial, if not intractable, we explore the issue starting from a simple HMM as shown in Figure 3.5, where the observation process contains only two states, and the underlying model is an illness-death model. When an individual is in either state 1 or 2 of the underlying model, we can only know that he/she is not in the absorbing state.

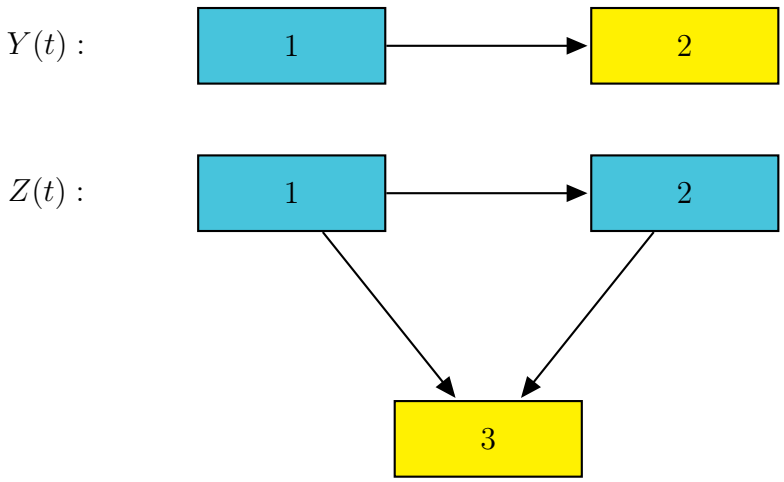


Figure 3.5: A simple hidden Markov model with a two-state observation process and a three-state underlying process

The emission probability matrix for such an HMM is

$$\mathrm{E} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The intensity matrix for the underlying model is

$$\Lambda = \begin{bmatrix} -h_1 & \lambda_{12} & \lambda_{13} \\ 0 & -\lambda_{23} & \lambda_{23} \\ 0 & 0 & 0 \end{bmatrix},$$

where $h_1 = \lambda_{12} + \lambda_{13}$. The closed forms of the elements of the transition probability matrix,

obtained by solving the Kolmogorov forward differential equation, are the following:

$$\begin{aligned} p_{11}(s,t) &= e^{-h_1(t-s)}, \\ p_{12}(s,t) &= \begin{cases} -\frac{\lambda_{12}}{h_1-\lambda_{23}}\left[e^{-h_1(t-s)} - e^{-\lambda_{23}(t-s)}\right] & \text{if } h_1 \neq \lambda_{23} \\ \lambda_{12}(t-s)\, e^{-h_1(t-s)} & \text{if } h_1 = \lambda_{23}, \end{cases} \\ p_{22}(s,t) &= e^{-\lambda_{23}(t-s)}, \end{aligned}$$

$p_{21} = p_{31} = p_{32} = 0$, and $p_{i3} = 1 - \sum_{j=1}^{2} p_{ij}, i = 1, 2, 3$. Suppose that all participants were recruited at time 0, let $\pi_i$ denote the probability for being in state $i$ at time 0, and $\pi_3 = 0$ and $\pi_2 = 1 - \pi_1$. The vector of unknown parameters is thus $\Theta = (\pi_1, \lambda_{12}, \lambda_{13}, \lambda_{23})$. Let $t$ be the death or censoring time whichever is less, and $\delta = 1$ if $Y(t) = 2$, and $\delta = 0$ otherwise. If $h_1 \neq \lambda_{23}$, the likelihood function is

$$\begin{aligned} \mathcal{L}(\Theta) &= \pi_1 p_{11}(0,t)\lambda_{13}^{\delta} + \pi_1 p_{12}(0,t)\lambda_{23}^{\delta} + (1-\pi_1)\, p_{22}(0,t)\lambda_{23}^{\delta} \\ &= \frac{\pi_1(\lambda_{13} - \lambda_{23})}{h_1 - \lambda_{23}}\left(h_1^{\delta} e^{-h_1 t} - \lambda_{23}^{\delta} e^{-\lambda_{23} t}\right) + \lambda_{23}^{\delta} e^{-\lambda_{23} t}, \end{aligned} \tag{3.20}$$

where we only have three effective parameters: $h_1 = \lambda_{12} + \lambda_{13}$, $\lambda_{23}$ and $(\pi_1(\lambda_{13} - \lambda_{23}))/(h_1 - \lambda_{23})$. Consider another set of values for the parameters $\Theta^* = (\pi_1^*, \lambda_{12}^*, \lambda_{13}^*, \lambda_{23}^*)$, where for some $0 < \rho < 1/\pi_1$,

$$\begin{aligned} \pi_1^* &= \rho\pi_1 \\ \lambda_{13}^* &= \frac{\lambda_{13} - (1-\rho)\lambda_{23}}{\rho} \\ \lambda_{12}^* &= \lambda_{12} + \lambda_{13} - \lambda_{13}^* \\ \lambda_{23}^* &= \lambda_{23}. \end{aligned}$$

It's obvious that $\mathcal{L}(\Theta) = \mathcal{L}(\Theta^*)$. If $h_1 = \lambda_{23}$, the likelihood function is

$$\mathcal{L}(\Theta) = e^{-h_1 t}\left[(1 + \pi_1\lambda_{12} t)\, h_1^{\delta} - \delta\pi_1\lambda_{12}\right]. \tag{3.21}$$

That is, for any $0 < \rho < 1/\pi_1$, the likelihood function gives the same value when $\pi_1^* = \rho\pi_1$,

$\lambda_{12}^* = \lambda_{12}/\rho$ and $\lambda_{12}^* = \lambda_{12} + \lambda_{13} - \lambda_{13}^*$.

As an extension to the HMM, the proposed mixture HMM has a more complicated structure, therefore the problems are more likely to appear. Consider a simplified scenario where all of the participants were followed from time 0 until their deaths. Assume that all of the participants started in the disease-free state, with or without pathology, and were diagnosed with dementia before they died. Let $a_1$ and $a_2$ be the time at the first and the last assessment time that the participant is in the disease-free state, i.e. $a_1 = \min\{t_k; Y(t_k) = 1\}$, $a_2 = \max\{t_k; Y(t_k) = 1\}$, $k = 0, 1, \ldots, K$. Similarly, let $a_3 = \min\{t_k; Y(t_k) = 2\}$, $a_4 = \max\{t_k; Y(t_k) = 2\}$, and $a_5 = t_D$. Assume that $Z^{(1)}(t)$ and $Z^{(2)}(t)$ are time-homogeneous Markov models and let $a_{uv} = a_v - a_u$ for $u < v, u, v \in \{1, 2, \ldots, 5\}$. If $\lambda_{12}^{(1)} + \lambda_{13}^{(1)} \neq \lambda_{23}^{(1)}$, $\lambda_{12}^{(2)} + \lambda_{14}^{(2)} \neq \lambda_{23}^{(2)} + \lambda_{25}^{(2)} \neq \lambda_{35}^{(2)}$, the likelihood function (3.18) can be written as

$$
\begin{aligned}
\mathcal{L} \;=\; & I_1\psi_1\pi_1^{(1)}\lambda_{12}^{(1)}\lambda_{23}^{(1)}\frac{e^{-\left(\lambda_{12}^{(1)}+\lambda_{13}^{(1)}\right)a_{12}-\lambda_{23}^{(1)}(a_{23}+a_{35})} - e^{-\left(\lambda_{12}^{(1)}+\lambda_{13}^{(1)}\right)(a_{12}+a_{23})-\lambda_{23}^{(1)}a_{35}}}{\lambda_{12}^{(1)}+\lambda_{13}^{(1)}-\lambda_{23}^{(1)}} \\
& + I_2\psi_2\left\{\pi_1^{(2)}\lambda_{12}^{(2)}\lambda_{23}^{(2)}\lambda_{35}^{(2)}\left(\frac{e^{-\left(\lambda_{12}^{(2)}+\lambda_{14}^{(2)}\right)a_{12}-\lambda_{35}^{(2)}(a_{23}+a_{35})}}{\left(\lambda_{12}^{(2)}+\lambda_{14}^{(2)}-\lambda_{35}^{(2)}\right)\left(\lambda_{23}^{(2)}+\lambda_{25}^{(2)}-\lambda_{35}^{(2)}\right)}\right.\right. \\
& + \frac{e^{-\left(\lambda_{12}^{(2)}+\lambda_{14}^{(2)}\right)(a_{12}+a_{23})-\lambda_{35}^{(2)}a_{35}}}{\left(\lambda_{12}^{(2)}+\lambda_{14}^{(2)}-\lambda_{23}^{(2)}-\lambda_{25}^{(2)}\right)\left(\lambda_{12}^{(2)}+\lambda_{13}^{(2)}-\lambda_{35}^{(2)}\right)} \\
& \left.\left. - \frac{e^{-\left(\lambda_{12}^{(2)}+\lambda_{14}^{(2)}\right)a_{12}-\lambda_{35}^{(2)}(a_{23}+a_{35})} + e^{-\left(\lambda_{23}^{(2)}+\lambda_{25}^{(2)}\right)(a_{12}+a_{23})-\lambda_{35}^{(2)}a_{35}} - e^{-\left(\lambda_{23}^{(2)}+\lambda_{25}^{(2)}\right)a_{12}-\lambda_{35}^{(2)}(a_{23}+a_{35})}}{\left(\lambda_{12}^{(2)}+\lambda_{14}^{(2)}-\lambda_{23}^{(2)}-\lambda_{25}^{(2)}\right)\left(\lambda_{23}^{(2)}+\lambda_{25}^{(2)}-\lambda_{35}^{(2)}\right)}\right) \\
& - \pi_2^{(2)}\lambda_{23}^{(2)}\lambda_{35}^{(2)}\frac{e^{-\left(\lambda_{23}^{(2)}+\lambda_{25}^{(2)}\right)(a_{12}+a_{23})-\lambda_{35}^{(2)}a_{35}} - e^{-\left(\lambda_{23}^{(2)}+\lambda_{25}^{(2)}\right)a_{12}-\lambda_{35}^{(2)}(a_{23}+a_{35})}}{\lambda_{23}^{(2)}+\lambda_{25}^{(2)}-\lambda_{35}^{(2)}}\right\}, \qquad (3.22)
\end{aligned}
$$

where $\psi_2 = 1 - \psi_1$ and

$$\begin{aligned} I_1 &= 1 - \mathbb{I}\{\Delta = 1, X = 1\} = \begin{cases} 1 & \text{if pathology result is negative or missing,} \\ 0 & \text{otherwise;} \end{cases} \\ I_2 &= 1 - \mathbb{I}\{\Delta = 1, X = 0\} = \begin{cases} 1 & \text{if pathology result is positive or missing,} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.23}$$

In (3.22), the vector of unknown parameters is

$$\Theta = \left( \psi_1,\ \pi_1^{(1)},\ \lambda_{12}^{(1)},\ \lambda_{13}^{(1)},\ \lambda_{23}^{(1)},\ \pi_1^{(2)},\ \pi_2^{(2)},\ \lambda_{12}^{(2)},\ \lambda_{14}^{(2)},\ \lambda_{23}^{(2)},\ \lambda_{25}^{(2)},\ \lambda_{35}^{(2)} \right).$$

It's obvious that we are able to identify $\lambda_{12}^{(1)} + \lambda_{23}^{(1)}$, $\lambda_{23}^{(1)}$, $\lambda_{12}^{(2)} + \lambda_{14}^{(2)}$, $\lambda_{23}^{(2)} + \lambda_{25}^{(2)}$, $\lambda_{35}^{(2)}$, and $\psi_1$. Consider some $\rho_1 \in (0, 1/\pi_1^{(1)})$ and $\rho_2 \in (0, 1/(\pi_2^{(1)} + \pi_2^{(2)}))$; it can be shown that the following parameter vector gives the same likelihood as $\Theta$ does.

$$\begin{aligned} \Theta^* = \Big( & \psi_1,\ \rho_1\pi_1^{(1)},\ \tfrac{\lambda_{12}^{(1)}}{\rho_1},\ \left(1 - \tfrac{1}{\rho_1}\right)\lambda_{12}^{(1)} + \lambda_{13}^{(1)},\ \lambda_{23}^{(1)}, \\ & \rho_2\pi_1^{(2)},\ \rho_2\pi_2^{(2)},\ \lambda_{12}^{(2)},\ \lambda_{14}^{(2)},\ \tfrac{\lambda_{23}^{(2)}}{\rho_2},\ \left(1 - \tfrac{1}{\rho_{23}}\right)\lambda_{23}^{(2)} + \lambda_{25}^{(2)},\ \lambda_{35}^{(2)} \Big) \end{aligned} \tag{3.24}$$

Therefore there is an identifiability issue in this case. However, if we have both participants who were diagnosed with dementia and participants who were never observed in the dementia state before passing away, we would be able to identify all of the parameters. A complete list of closed form expressions of the likelihood function for different observed data scenarios is given in Appendix A.1.

Even if all of the parameters are identifiable according to the form of the likelihood function, we still cannot guarantee to produce good estimates from numerically maximizing the likelihood function. The complicated form of the likelihood function brings difficulties in the estimation of the parameters. One reason behind the difficulties in getting the MLE is the existence of local maxima in the likelihood function, which may cause the maximizing procedure to be stuck at a point away from the MLE. Another reason is the "weak identifiability", which occurs when the likelihood has "flat" regions around the

Table 3.5: Comparison of the choices of constraints based on simulation studies with 700 replications and sample size 500

| | | $S_0$: No cons. | | $S_1$: $\frac{\widehat{\pi}_2^{(2)}}{\widehat{\pi}_1^{(2)}} = \frac{\pi_2^{(2)}}{\pi_1^{(2)}}$ | | $S_2$: $\frac{\widehat{\lambda}_{14}^{(2)}}{\widehat{\lambda}_{13}^{(1)}} = \frac{\lambda_{14}^{(2)}}{\lambda_{13}^{(1)}}$ | | $S_3$: $S_1$&$S_2$ | | $S_4$: $\frac{\widehat{\lambda}_{12}^{(2)}}{\widehat{\lambda}_{14}^{(2)}} = \frac{\lambda_{25}^{(2)}}{\lambda_{14}^{(2)}}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | TRUE | EST | SE | EST | SE | EST | SE | EST | SE | EST | SE |
| | | | | | | *Transition Intensities of* $Z^{(1)}(t)$ | | | | | |
| $\lambda_{12}^{(1)}$ | 2.383 | 2.441 | 0.713 | 3.166 | 1.029 | 2.269 | 0.356 | 3.595 | 0.812 | 2.832 | 0.599 |
| $\lambda_{13}^{(1)}$ | 1.191 | 1.272 | 0.275 | 0.947 | 0.555 | 1.275 | 0.257 | 0.640 | 0.382 | 1.214 | 0.319 |
| $\lambda_{23}^{(1)}$ | 1.787 | 2.157 | 0.323 | 2.169 | 0.323 | 2.148 | 0.320 | 2.160 | 0.321 | 2.179 | 0.325 |
| | | | | | | *Transition Intensities of* $Z^{(2)}(t)$ | | | | | |
| $\lambda_{12}^{(2)}$ | 1.802 | 0.000 | 0.005 | 1.447 | 0.424 | 0.000 | 0.000 | 1.267 | 0.204 | 0.000 | 0.004 |
| $\lambda_{14}^{(2)}$ | 0.819 | 0.274 | 0.835 | 0.610 | 0.466 | . | . | . | . | 0.803 | 0.127 |
| $\lambda_{23}^{(2)}$ | 2.457 | 1.386 | 0.254 | 2.474 | 0.610 | 1.390 | 0.256 | 2.296 | 0.480 | 1.346 | 0.245 |
| $\lambda_{25}^{(2)}$ | 1.474 | 1.555 | 0.290 | 2.665 | 0.674 | 1.577 | 0.283 | 2.455 | 0.503 | . | . |
| $\lambda_{35}^{(2)}$ | 2.047 | 1.449 | 0.226 | 1.437 | 0.221 | 1.454 | 0.227 | 1.438 | 0.221 | 1.435 | 0.220 |
| | | | | | | *Other Parameters* | | | | | |
| $\pi_1^{(1)}$ | 0.700 | 0.669 | 0.312 | 0.625 | 0.329 | 0.677 | 0.249 | 0.599 | 0.212 | 0.649 | 0.243 |
| $\pi_1^{(2)}$ | 0.400 | 0.076 | 0.076 | 0.419 | 0.051 | 0.057 | 0.027 | 0.426 | 0.039 | 0.111 | 0.058 |
| $\pi_2^{(2)}$ | 0.300 | 0.627 | 0.311 | . | . | 0.639 | 0.234 | . | . | 0.607 | 0.245 |
| $\psi_1$ | 0.500 | 0.495 | 0.161 | 0.437 | 0.153 | 0.508 | 0.103 | 0.409 | 0.087 | 0.466 | 0.102 |

maximum value of likelihood, so that the MLE yields multiple different but close estimates. Berzofsky and Biemer (2012) suggest that this is a common issue for latent class models.

Literature suggests that more frequent observations will make the problem less likely to appear (Titman and Sharples, 2010). However, increasing observation frequency can be expensive and is hard to achieve if the issue was found after the data collection is done. In this case, introducing constraints on the transition intensities to reduce the dimension of the parameter space may help.

In our simulation study for the model in Figure 3.2, we dealt with the issue with a dimension reduction approach. Table 3.5 gives several options of constraints and their performance. These options include: scenario 0 - directly maximizing the likelihood without

constraint; scenario 1 (S1) - controlling the ratio of being in pathology state versus healthy state for disease type 2; scenario 2 (S2) - controlling the ratio of intensities from disease-free to death of the two disease types; scenario 3 (S3) - controlling the previous two at the same time; scenario 4 (S4) - controlling the ratio of intensities from dementia to death of the two disease types. The estimates and the model-based standard errors in Table 3.5 are averaged over 100 replications. Without imposing any constraint, the MLE and standard error for $\lambda_{12}^{(2)}$ gave zero values, suggesting the presence of potential issues. Scenarios S1 and S3 produced valid estimates for $\lambda_{12}^{(2)}$ while S1 required less reduction in dimension of the parameter. This is the reason why we proceeded with scenario S1. However, the choice of the constraint depends on the structure of the model and values of the true intensities. Standard solutions that apply to the general identifiability and estimability issues in mixture HMMs are still under investigation.

For the Nun Study case, we put a constraint on the initial probability for being in the pathology state at age 75 to reduce the parameter dimension. However, the way to deal with the estimability issue is not unique. For example, information regarding the relationship between the transition intensities may be available in the literature. There may be information on the comparison between time-to-dementia for a no-AD-pathology participant versus a participant with pathology, or the death time difference between an individual with dementia and a healthy individual. This kind of information can be either a definite number or an approximate interval.

## 3.6 A Bayesian Approach for Parameter Estimation

The Bayesian method has became a popular tool for making statistical inference using complex models due to the convenience in numerical calculation and the ability to incorporate knowledge on the parameters. It usually involves assignment of prior distributions to the parameters and, possibly, a set of hyperparameters of the prior distributions. The Bayesian inference is made based on the posterior distributions of the parameters. The posteriors are derived based on the priors and the conditional distribution of the observed data given the values of the parameters, thus are proportional to the likelihood function.

The standard point estimator of the value of a parameter is the mean of the posterior for that parameter, given the data.

In this section, we consider Bayesian inference for the proposed mixture HMM and explore if incorporating prior informations on certain parameters can help with the identifiability and estimability issues. Simulations are conduced by using the same model and parameters configurations as in Section 3.4. Results are based on 2000 iterations (excluding the burning stage) and sample size 5000. The numerical calculations use the Gibbs Sampler conducted by JAGS. We use two criteria to evaluate the convergence of the Gibbs Sampler, the effective number of iterations $N_{\text{eff}}$ and another commonly used criterion $\widehat{R}$ (Gelman and Rubin, 1992). Usually, we expect $N_{\text{eff}} \geq 30$ and $\widehat{R} \approx 1$ if the algorithm converges successfully.

Given the parameters of interest $\Theta = \left\{\psi_m, \pi^{(m)}, \lambda^{(m)}, m \in \{1,2\}\right\}$, consider three scenarios for the prior distributions. Details of the prior distribution and the analysis results are given below.

For scenario 1, we use diffuse prior distributions for all of the parameters, which are in forms as the following:

$$\begin{aligned}
\psi_1 &\sim \text{Uniform}\,(0,1) \\
\pi_1^{(1)} &\sim \text{Uniform}\,(0,1) \\
\pi_1^{(2)} &\sim \text{Uniform}\,(0,1) \\
\pi_2^{(2)} \mid \pi_1^{(2)} &\sim \text{Uniform}(0, 1-\pi_1^{(2)}) \\
\log \lambda_{ij}^{(m)} &\sim \text{Normal}\,(0, 10000)\,.
\end{aligned}$$

Table 3.6 gives results for scenario 1. When we have no prior knowledge on any parameter, the algorithm gives posteriors that are almost fixed at zero for $\lambda_{12}^{(2)}$.

Scenario 2 shows an ideal case where the prior distributions of transition intensities have means at the true values of the parameters. The prior distributions for scenario 2 are

the following:

$$\begin{aligned}
\psi_1 &\sim \text{Uniform}\,(0,1) \\
\pi_1^{(1)} &\sim \text{Uniform}\,(0,1) \\
\pi_1^{(2)} &\sim \text{Uniform}\,(0,1) \\
\pi_2^{(2)} \mid \pi_1^{(2)} &\sim \text{Uniform}(0, 1-\pi_1^{(2)}) \\
\log \lambda_{ij}^{(m)} &\sim \text{Normal}\left(\log \lambda_{ij}^{(m)}, V\right) \\
V &\sim \text{Uniform}\,(0.2, 0.3)\,.
\end{aligned}$$

When using informative priors, the results become much better; see Table 3.6, where valid estimates of parameters and standard errors are produced and most of the $N_{eff}$ and $\widehat{R}$ are acceptable.

One may wonder, as the problem is most likely with the estimation of $\lambda_{12}^{(2)}$ due to the fact that onset time of AD pathology is unobservable and only current status information is available, would it be helpful to only provide prior information on this intensity? Thus we tried scenario 3, where we put informative priors for $\lambda_{14}^{(2)}$ and $\lambda_{12}^{(2)}/\lambda_{14}^{(2)}$ and diffuse priors for the rest of the parameters. Instead of directly putting prior information on $\lambda_{12}^{(2)}$, we use prior information on the transition intensity for the disease-free death, $\lambda_{14}^{(2)}$, and the ratio between developing AD pathology versus death. This is because that we cannot observe transitions from the disease-free state to the pathology state, while it is might be possible to get prior information on the disease-free mortality rate, and the ratio of portions of individuals who died before and after the pathology developed. Another reason of putting prior information on intensities for transitions $1 \to 2$ and $1 \to 4$ is that these transitions are never observed, which is likely to cause identifiability and estimability issues. The prior

distributions for scenario 3 are the following:

$$
\begin{aligned}
\psi_1 &\sim \text{Uniform}\,(0,1) \\
\pi_1^{(1)} &\sim \text{Uniform}\,(0,1) \\
\pi_1^{(2)} &\sim \text{Uniform}\,(0,1) \\
\pi_2^{(2)} \mid \pi_1^{(2)} &\sim \text{Uniform}(0, 1-\pi_1^{(2)}) \\
\log \lambda_{ij}^{(m)} &\sim \text{Normal}\,(0, 10000)\,, \text{excludes } \log \lambda_{12}^{(2)}, \log \lambda_{14}^{(2)} \\
\log \lambda_{ij}^{(m)} &\sim \text{Normal}\left(\log \lambda_{ij}^{(m)}, V\right) \\
\frac{\lambda_{12}^{(2)}}{\lambda_{14}^{(2)}} &\sim \text{Normal}\left(\frac{\lambda_{12}^{(2)}}{\lambda_{14}^{(2)}}, V\right) \\
V &\sim \text{Uniform}\,(0.2, 0.3)\,.
\end{aligned}
$$

Table 3.6 shows that the informative prior we chose for $\lambda_{12}^{(2)}$ and $\lambda_{14}^{(2)}$ does help with the estimation of $\lambda_{12}^{(2)}$. However, comparing the results using type 1 and type 3 priors, only using informative priors for $\lambda_{12}^{(2)}$ and $\lambda_{14}^{(2)}$ causes problems in posteriors of $\lambda_{23}^{(2)}$ and $\lambda_{25}^{(2)}$. Therefore it is not clear if utilizing auxiliary information on a problematic parameter will solve the issue. Instead, one needs to find the solution case by case.

Table 3.6: Bayesian analysis based on a simulated data set with sample size 5000 using different types of priors. Type 1 - using diffuse prior distributions; Type 2 - using informative prior distributions which have means at the true values of the parameters; Type 3 - using informative prior distributions for $\lambda_{14}^{(2)}$ and $\lambda_{12}^{(2)}/\lambda_{14}^{(2)}$ which have means at the true values and diffuse priors for the other parameters. $q_p$: the $q^{\text{th}}$ quantile of the posterior distributions. $N_{\text{eff}} \geq 30$ and $\widehat{R} \approx 1$ indicate that the algorithm converges successfully.

| | | Posterior Distribution | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | TRUE | MEAN | SD | $q_{2.5\%}$ | $q_{25\%}$ | $q_{50\%}$ | $q_{75\%}$ | $q_{97.5\%}$ | $\widehat{R}$ | $N_{\text{eff}}$ |
| *Using Type 1 Prior Distributions* | | | | | | | | | | |
| $\lambda_{12}^{(1)}$ | 2.383 | 3.594 | 0.584 | 2.595 | 3.142 | 3.532 | 3.966 | 4.846 | 1.001 | 2000 |
| $\lambda_{13}^{(1)}$ | 1.191 | 0.010 | 0.064 | 0.000 | 0.000 | 0.000 | 0.000 | 0.088 | 1.001 | 1800 |
| $\lambda_{23}^{(1)}$ | 1.787 | 2.166 | 0.228 | 1.754 | 2.003 | 2.150 | 2.316 | 2.640 | 1.007 | 660 |
| $\lambda_{12}^{(2)}$ | 1.802 | 0.000 | 0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 2.260 | 3 |
| $\lambda_{14}^{(2)}$ | 0.819 | 3.802 | 0.634 | 2.803 | 3.333 | 3.751 | 4.194 | 5.204 | 1.002 | 930 |
| $\lambda_{23}^{(2)}$ | 2.457 | 1.309 | 0.130 | 1.074 | 1.220 | 1.301 | 1.393 | 1.578 | 1.010 | 160 |
| $\lambda_{25}^{(2)}$ | 1.474 | 1.115 | 0.129 | 0.881 | 1.021 | 1.111 | 1.200 | 1.387 | 1.006 | 2000 |
| $\lambda_{35}^{(2)}$ | 2.047 | 1.560 | 0.168 | 1.266 | 1.446 | 1.549 | 1.661 | 1.954 | 1.028 | 620 |
| $\pi_1^{(1)}$ | 0.700 | 0.608 | 0.034 | 0.538 | 0.587 | 0.608 | 0.632 | 0.670 | 1.001 | 2000 |
| $\pi_1^{(2)}$ | 0.400 | 0.194 | 0.024 | 0.148 | 0.178 | 0.193 | 0.209 | 0.242 | 1.012 | 2000 |
| $\pi_2^{(2)}$ | 0.300 | 0.573 | 0.025 | 0.520 | 0.558 | 0.574 | 0.590 | 0.621 | 1.015 | 2000 |
| $\psi_1$ | 0.500 | 0.368 | 0.025 | 0.320 | 0.350 | 0.368 | 0.384 | 0.421 | 1.006 | 2000 |
| *Using Type 2 Prior Distributions* | | | | | | | | | | |
| $\lambda_{12}^{(1)}$ | 2.383 | 2.954 | 0.498 | 2.106 | 2.601 | 2.903 | 3.241 | 4.069 | 1.002 | 1100 |
| $\lambda_{13}^{(1)}$ | 1.191 | 0.981 | 0.272 | 0.453 | 0.796 | 0.983 | 1.166 | 1.533 | 1.014 | 140 |
| $\lambda_{23}^{(1)}$ | 1.787 | 2.085 | 0.244 | 1.643 | 1.907 | 2.076 | 2.255 | 2.577 | 1.013 | 130 |
| $\lambda_{12}^{(2)}$ | 1.802 | 1.968 | 0.424 | 1.252 | 1.675 | 1.931 | 2.223 | 2.881 | 1.090 | 28 |
| $\lambda_{14}^{(2)}$ | 0.819 | 0.711 | 0.249 | 0.337 | 0.532 | 0.675 | 0.851 | 1.235 | 1.035 | 2000 |
| $\lambda_{23}^{(2)}$ | 2.457 | 2.407 | 0.585 | 1.478 | 1.961 | 2.351 | 2.739 | 3.746 | 1.094 | 26 |
| $\lambda_{25}^{(2)}$ | 1.474 | 1.958 | 0.490 | 1.134 | 1.583 | 1.892 | 2.287 | 2.994 | 1.100 | 26 |
| $\lambda_{35}^{(2)}$ | 2.047 | 1.640 | 0.197 | 1.326 | 1.503 | 1.605 | 1.756 | 2.075 | 1.012 | 140 |
| $\pi_1^{(1)}$ | 0.700 | 0.661 | 0.033 | 0.595 | 0.640 | 0.662 | 0.684 | 0.722 | 1.010 | 270 |
| $\pi_1^{(2)}$ | 0.400 | 0.393 | 0.111 | 0.139 | 0.328 | 0.409 | 0.471 | 0.579 | 1.312 | 14 |
| $\pi_2^{(2)}$ | 0.300 | 0.352 | 0.105 | 0.177 | 0.276 | 0.336 | 0.413 | 0.579 | 1.149 | 38 |
| $\psi_1$ | 0.500 | 0.433 | 0.033 | 0.372 | 0.410 | 0.433 | 0.454 | 0.500 | 1.001 | 1600 |
| *Using Type 3 Prior Distributions* | | | | | | | | | | |
| $\lambda_{12}^{(1)}$ | 2.383 | 3.662 | 0.692 | 2.475 | 3.169 | 3.611 | 4.108 | 5.150 | 1.033 | 66 |
| $\lambda_{13}^{(1)}$ | 1.191 | 0.011 | 0.071 | 0.000 | 0.000 | 0.000 | 0.000 | 0.136 | 1.001 | 2000 |
| $\lambda_{23}^{(1)}$ | 1.787 | 2.098 | 0.247 | 1.670 | 1.923 | 2.083 | 2.245 | 2.647 | 1.011 | 150 |
| $\lambda_{12}^{(2)}$ | 1.802 | 1.908 | 0.226 | 1.442 | 1.781 | 1.891 | 2.026 | 2.409 | 1.162 | 1200 |
| $\lambda_{14}^{(2)}$ | 0.819 | 0.898 | 0.253 | 0.589 | 0.726 | 0.834 | 0.992 | 1.572 | 2.569 | 3 |
| $\lambda_{23}^{(2)}$ | 2.457 | 137.810 | 215.694 | 1.921 | 4.201 | 19.187 | 203.291 | 755.637 | 6.561 | 2 |
| $\lambda_{25}^{(2)}$ | 1.474 | 114.434 | 177.025 | 1.555 | 3.741 | 16.432 | 177.552 | 620.473 | 6.499 | 2 |
| $\lambda_{35}^{(2)}$ | 2.047 | 1.615 | 0.195 | 1.240 | 1.490 | 1.601 | 1.717 | 2.072 | 1.007 | 230 |
| $\pi_1^{(1)}$ | 0.700 | 0.603 | 0.034 | 0.533 | 0.581 | 0.605 | 0.626 | 0.663 | 1.006 | 2000 |
| $\pi_1^{(2)}$ | 0.400 | 0.680 | 0.109 | 0.416 | 0.622 | 0.726 | 0.763 | 0.795 | 2.712 | 3 |
| $\pi_2^{(2)}$ | 0.300 | 0.090 | 0.106 | 0.000 | 0.005 | 0.036 | 0.150 | 0.349 | 4.663 | 2 |
| $\psi_1$ | 0.500 | 0.369 | 0.027 | 0.316 | 0.350 | 0.369 | 0.387 | 0.424 | 1.021 | 78 |

## 3.7 Discussion

In this chapter, we proposed a mixture hidden Markov model which provides the ability to model non-Markov transition times while utilizing the good properties of Markov models. Comparing to the traditional HMM, the mixture HMM is particularly appealing when the underlying disease has multiple types and when relaxation of the Markov assumption on the observation model is needed. The proposed method is also able to incorporate potential information for the mixture membership to improve the efficiency and the estimability.

An advantage of the mixture hidden Markov models is that the estimates can be interpreted for cohorts with either single-type of disease or mixture-type of disease. In our application to the disease process of dementia, the model was able to show the difference in prevalence of the two types of disease. However, one needs to be careful when relating the estimates for a single-type cohort to the targeted disease type, because the potential identifiability and estimability issues may make these results less interpretable.

We have shown that even the simplest HMM suffers from the identifiability and estimability issues. This suggests that strategies to identify and contain these issues are essential to getting good estimates in mixture HMM. However, general procedures on how to do these are still under development. In this chapter, we used both analytical and empirical ways to identify the existence of these issues, and reduced the dimension of the unknown parameter by putting constraints to solve these issues. We also conducted Bayesian analyses to investigate the usage of auxiliary information on the parameter and constraints on the parameter. Results suggested that informative priors are helpful, while how to choose which parameter should have an informative prior depends on the structure of the model and the sample, and is thus not very straightforward. There are other solutions to these problems available from the literature (Titman and Sharples, 2010) including increasing the number of observations.

In conclusion, mixture HMM provides a good tool for modeling a disease process where multiple underlying disease types share common symptoms. It relaxes the Markov assumption and allows individuals to have different incidence rates of disease stages according to their disease types. Discussion on future research directions will be given in Chapter 5.

# Chapter 4

# Response Dependent Sampling in Multistate Models

## 4.1 Introduction

The Nun Study is a longitudinal study aiming to understand the disease progression and potential risk factors of AD. Participants of the study are 678 American Roman Catholic sisters who are members of the congregation of the School Sisters of Notre Dame. These participants were born between 1886 and 1916, and their ages at time of recruitment were 75 to 106.

The observations include the year of birth and approximately annual measurements of cognitive status for the participants from 1991 (the start of the study) until death or the end of follow up in 2003. Covariates include the presence of the allele $\epsilon 4$ in the apolipoprotein E gene ($APOE - \epsilon 4$), which is a commonly known risk factor for Alzheimer' s disease, and intellectual factors such as educational level, high school course grades, and number of languages spoken.

At each assessment time, the cognitive stages are categorized into three levels: normal cognition (NC), mild cognitive impairment (MCI) and dementia. We use a four-state

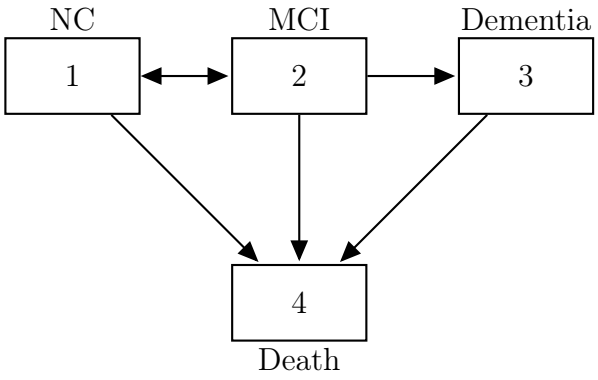


Figure 4.1: A four-state stochastic process for joint modeling of the transitions among normal cognition (NC), mild cognitive impairment (MCI), dementia, and death

model to represent the disease process for dementia, see Figure 4.1, where state 1, 2, 3 and 4 represent normal cognition, mild cognitive impairment, dementia and death correspondingly. The transitions include forward progression from normal cognition to mild cognitive impairment and then to dementia, a backward transition from mild cognitive impairment to normal cognition, as well as death that can happen at any disease level as a competing event. While the exact death time will be recorded, the exact transition times into other states are interval censored under the intermittent observation scheme. As AD is often a senile disease, we use age 75 as the time origin. The time scale used in this chapter is time since turning 75 (i.e., age - 75).

In many longitudinal studies, the cohort is selected conditioning on subjects alive and in one of states 1, 2 or 3 at the study entry. For example in the Nun Study, only religious sisters from the original birth cohort that were still alive in 1991 were eligible to be recruited into the study. This will lead to left-truncation in the sample. In a longitudinal study of bloodstream infection (Vakulenko-Lagun and Mandel, 2016), a cross-sectional sample of intensive care unit patients who lived long enough to be present on the sampling day were followed until discharge or death. Andersen (1988) analyzed data from a study on incidence of nephropathy in insulin dependent diabetes, the study population consists of patients with diabetes and referred to a special hospital, and some of them already had diabetic nephropathy at the first admission. A similar situation appears in other prospective cohort

studies of disease processes such as HIV infection (Copas and Farewell, 2001) and dementia and AD (Commenges et al., 2004). Ignoring such sampling conditions will lead to bias in multistate analysis. One can construct likelihood conditioning on the baseline information, e.g., duration from time origin to the study entry and the state occupation at study entry. However, such a conditioning approach will generally result in a loss of information relative to that contained in the unconditional likelihood. The efficiency depends on the amount of information which is lost by ignoring the contribution of the baseline states and left-truncation time; parameters associated with certain transitions may not be identifiable when there is not enough information to recover as a consequence. We propose a weighted method that makes use of the state occupation at recruitment given the sampling condition and initial probabilities. As the disease process of dementia is widely studied, information regarding the age-specific prevalence for each cognitive stage and dementia can be found in the literature. The population prevalences of cognitive stages and dementia are used as the initial probabilities.

The remainder of this chapter is structured as follows. In Section 4.2, we define notation and introduce methods for multistate Markov models under intermittent observation. Then we develop a weighted likelihood which accommodates auxiliary population prevalence information to adjust for the left-truncation. In Section 4.3, we illustrate the application of the proposed method in the context of the Nun Study. In Section 4.4, we conduct simulation studies to illustrate the performance of the proposed method. we produce estimates using a Bayesian approach in Section 4.5 and discuss the sensitivity to the choice of the prior distribution. Concluding remarks and suggestions for future work are given in Section 4.6.

## 4.2 Methods

### 4.2.1 Multistate Models for Disease Process Data

The multistate model is a widely used tool for disease process data. Structures of such models contain multiple states which correspond to different stages of diseases and transi-

tions between these states. For example, let $\{Z(t); t \geq 0\}$ be a continuous time multistate stochastic process with state space $\mathcal{J} = \{1, 2, \ldots, J\}$. Let $\mathcal{H}(t) = \{Z(s); 0 \leq s \leq t\}$ be the history of state occupancies before time $t$; the transition intensity from state $i$ to state $j$ at time $t$ is

$$\lambda_{ij}(t \mid \mathcal{H}(t^-)) = \lim_{\Delta t \to 0} \frac{P\left(Z(t + \Delta t^-) = j | Z(t^-) = i, \mathcal{H}(t^-)\right)}{\Delta t}, \text{ for } i \neq j \text{ and } i, j \in \mathcal{J}. \quad (4.1)$$

One of the most widely used models for multistate processes is the continuous time Markov model which assumes the probability for being in a state at any future time $u > t$ given the current state occupancy $Z(t)$ is independent of transition history $\mathcal{H}(t^-)$. This is equivalent to assuming that

$$\lambda_{ij}(t \mid \mathcal{H}(t^-)) = \lambda_{ij}(t), \text{ for } i \neq j \text{ and } i, j \in \mathcal{J}. \quad (4.2)$$

Let $\Lambda(t)$ be the transition intensity matrix with the $(i, j)$ off-diagonal entries being $\lambda_{ij}(t)$, and the diagonal entries being $-h_i(t)$, where $h_i(t) = \sum_{j \neq i} \lambda_{ij}(t)$. Let P be the transition probability matrix where the $(i, j)$ entry is

$$p_{ij}(s, t) = P\left(Z(t) = j | Z(s) = i\right), \text{ for } 0 \leq s \leq t \text{ and } i, j \in \mathcal{J}. \quad (4.3)$$

Under the Markov assumption (4.2), the values of the transition probabilities can be obtained by solving the Kolmogorov forward differential equation

$$\frac{\partial}{\partial t} \mathrm{P}(s, t) \quad = \quad \mathrm{P}(s, t)\, \Lambda(t). \quad (4.4)$$

If we assume the intensities going from one state to the others are constants over time, i.e., $\Lambda(t) = \Lambda$, the solution to (4.4) is

$$\mathrm{P}(s, t) \quad = \quad e^{(t-s)\Lambda}. \quad (4.5)$$

This is the so-called time-homogeneous Markov model. When the intensity matrix $\Lambda$ has a relative simple structure, the closed forms of entries of (4.5) can be derived by matrix decomposition of $(t - s)\Lambda$. When $\Lambda$ is complicated, one can use a Taylor series to

approximate the value of the matrix exponential.

When the time-homogeneous assumption is violated, a piecewise constant intensity model provides a more flexible setting while retaining some convenience when calculating the transition probability matrix. Let $0 < c_1 < \ldots < c_R < \infty$ be pre-fixed cut-points. Without loss of generality let $c_0 = 0$ and $c_{R+1} = \infty$. Then the time intervals are $\mathcal{C}_r = [c_r, c_{r+1}), r = 0, 1, \ldots, R$. The transition intensity matrix for the $r^{\text{th}}$ interval is $\Lambda(t) = \Lambda_r$, $\forall t \in \mathcal{C}_r$. For $c_r < s < t < c_{r+1}$, the transition probability matrix is $\mathrm{P}(s,t) = e^{(t-s)\Lambda_r}, r = 0, 1, \ldots, R$. For $s \in \mathcal{C}_{r_s}$, $t \in \mathcal{C}_{r_t}$, the transition matrix can be calculated by recurrent use of the Chapman-Kolmogorov Equation, which gives

$$\mathrm{P}(s,t) = \prod_{r=r_s}^{r_t} \mathrm{P}(\max\{s, c_r\}, \min\{t, c_{r+1}\}). \tag{4.6}$$

When the research interests lie in investigating effects of risk factors, multiplicativ intensity models can be used. Suppose a vector of risk factors $X = (X_1, X_2, \ldots, X_P)$, and assume $X$ is invariant over time. Let $\Lambda_0$ and $\lambda_{ij0}$ be the baseline transition intensity matrix and its entries. The model for transition intensities is

$$\lambda_{ij}(t; X) = \lambda_{ij0}(t) e^{\beta_{ij}^{\mathrm{T}} X}, \tag{4.7}$$

where $\beta_{ij} = (\beta_{ij,1}, \beta_{ij,2}, \ldots, \beta_{ij,P})$ is the vector of log instantaneous rate ratios for transition from state $i$ to state $j$. Consider a piecewise-constant intensity model, let $\lambda_{ij,r}$ denote the intensity for $t \in \mathcal{C}_r$ for the participants with all of the covariates being 0, and let $\alpha_{ij,r} = \log(\lambda_{ijr}/\lambda_{ij0})$. The intensity model for piecewise-constant models with proportional hazard effects is

$$\lambda_{ij}(t; X) = \lambda_{ij0} e^{\sum_{p=1}^{P} X_p \beta_{ij,p} + \sum_{r=1}^{R} \mathbb{I}\{t \in \mathcal{C}_r\} \alpha_{ij,r}}. \tag{4.8}$$

### 4.2.2 Weighted Likelihood Incorporating the State at Sampling and Auxiliary Prevalence Information

In the Nun Study, the cognitive functions of the participants were assessed approximately annually. If death occurred during the study period, the exact time was observed. A mixed intermittent and continuous observation scheme is common in many prospective studies of chronic disease processes.

Suppose there are $K^*$ pre-fixed assessment times $0 \leq a_0 < a_1 < a_2 < \ldots < a_{K^*} = \tau$, where $a_0$ is the time of initial assessment in the working time scale (e.g., age at study entry - 75). Let $\tau$ denote the administrative right censoring time. Suppose the patients may drop out from the study at a random time $C$. Let $T_D$ denote the time to enter the terminal state (death). Denote the actual number of observations for a patient as $K = \max\{k; a_k \leq T^\dagger, k = 1, ..., K^*\}$, where $T^\dagger = \min\{T_D, C, \tau\}$ denotes the minimum of death or censoring time. Then the set of disease status observations for a patient is $\{Z(a_0), Z(a_1), \ldots, Z(a_K)\}$. The indicator of observing a death event is $\delta = \mathbb{I}\{\min\{\tau, C, ^* T_D\} = T_D\}$.

Condition on the left-truncation time $a_0$ and state at sampling $Z(a_0)$, the likelihood function for a subject with panel observation and exact death time is

$$\begin{aligned}
\mathcal{L}_0 &= P\left(Z(a_1) = z_1, ..., Z(a_K) = z_K, t^\dagger, \delta \mid Z(a_0) = z_0, a_0, X\right) \\
&= \prod_{k=1}^{K} \Bigg\{ P\left(Z(a_k) = z_k \mid Z(a_{k-1}) = z_{k-1}, X\right) \\
&\quad \times \left[ \sum_{j \neq 4} P\left(Z(t^\dagger) = j \mid Z(a_K) = z_K, X\right) \lambda_{j4}(t^\dagger, X) \right]^{\delta} \Bigg\}, \qquad (4.9)
\end{aligned}$$

The component in the brackets takes into account the unknown state occupancy right before death occurs. The conditional maximum likelihood estimates are obtained by maximizing formula (4.9). Such estimates can be applied to the cohort of patients who are alive at the time of recruitment.

In order to utilize the baseline state occupation information, we borrow auxiliary information of chronic disease processes to help us to consider the sampling condition, which

is $Z(a_0) \neq 4$. The condition is incorporated to the likelihood (4.9) in terms of weights. Suppose from the literature or previous studies, we know the prevalence for being in state $j$ at the time origin 0, i.e., age of 75, $j \neq 4$. This can be denoted in terms of the initial probability for being in state $j$, namely $\pi_j = P\left(Z(0) = j\right), j \neq 4$. Then we consider the probability for an individual to be eligible (i.e., alive) at the beginning of the study, and obtain the adjusted likelihood with form

$$\mathcal{L} \;=\; w\mathcal{L}_0, \tag{4.10}$$

where

$$\begin{aligned} w &= P\left(Z(a_0) = z_0 \mid Z(a_0) \neq 4, a_0\right) \\ &= \frac{\sum_{j=1}^{3} P\left(Z(a_0)|Z(0) = j, a_0\right)\pi_j}{\sum_{j=1}^{3} P\left(Z(a_0) \neq 4|Z(0) = j, a_0\}\pi_j\right.}. \end{aligned} \tag{4.11}$$

The MLE for transition intensities can be obtained by maximizing the weighted likelihood (4.10). Note that this method requires initial probabilities $\pi_j$'s to be known or estimated.

## 4.3 Results from the Nun Study Data

For the Nun Study data, consider the multistate model given in Figure 4.1. The intensity matrix for this model has the following structure:

$$\Lambda(t) \;=\; \begin{bmatrix} -h_1(t) & \lambda_{12}(t) & 0 & \lambda_{14}(t) \\ \lambda_{21}(t) & -h_2(t) & \lambda_{23}(t) & \lambda_{24}(t) \\ 0 & 0 & -h_3(t) & \lambda_{34}(t) \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4.12}$$

To apply the likelihood function (4.10), we obtain the prevalences $\pi_j, j = 1, ..., 4$ at the time origin from the literature. In Yesavage et al. (2002), the prevalences for mild cognitive impairment and dementia at age 75 for the general population are estimated as 0.2

Table 4.1: Estimated prevalence for dementia at 75 for the general population. NC: Normal cognition; MCI: Mild cognitive impairment.

| State | NC | MCI | Dementia |
|---|---|---|---|
| Prevalence | 0.68 | 0.20 | 0.12 |

and 0.12 respectively, see Table 4.1. This is equivalent to an estimate for $\pi$, such that $\pi = (0.68, 0.20, 0.12)$. Borrowing this information, we fit the model using the weighted likelihood $\mathcal{L}$ given in (4.10) and compare with the results from the conditional likelihood $\mathcal{L}_0$ given in (4.9).

We compared results obtained using the conditional likelihood (4.9) and the weighted likelihood (4.10). We considered two intensity models. Table 4.2 gives the results from a piecewise constant intensities model with breakpoint at age 90, where the intensities are

$$\lambda_{ij}(t) = \lambda_{ij0} e^{\alpha_{ij} \mathbb{I}\{t>90\}}, \tag{4.13}$$

and the parameter of interest is $\Theta = \{\lambda_{ij0}, \alpha_{ij}; i, j, \in \mathcal{J}\}$. $APOE-\epsilon4$ is a genotype that is known to be a risk factor for Alzheimer's disease. Let $X = 1$ if a participant has $APOE-\epsilon4$, and 0 otherwise. Table 4.2 also gives the results from a proportional hazard intensity model for $X$ with a breakpoint at age 90.

$$\lambda_{ij}(t; X) = \lambda_{ij0} e^{\alpha_{ij} \mathbb{I}\{t>90\} + \beta_{ij} X}, \tag{4.14}$$

where $\Theta = \{\lambda_{ij0}, \alpha_{ij}, \beta_{ij}; i, j, \in \mathcal{J}\}$. To help with the identifiability and estimability issues, we assume that the covariate does not have effects on transitions to the death state, i.e., $\beta_{14} = \beta_{24} = \beta_{34} = 0$. These constraints are based on results from analyses considering $APOE-\epsilon4$'s effect on transitions to death, which suggest that $APOE-\epsilon4$'s effect on mortality is not statistically significant. The standard errors in Table 4.2 are the square root of the model-based variances from the inverse of the negative of the information matrix. The gain in efficiency for a parameter $\theta$ is seen by the relative reduction in the

model-based variances after utilizing the auxiliary information, i.e.,

$$RV = \frac{\mathrm{Var}(\widetilde{\theta}) - \mathrm{Var}(\widehat{\theta})}{\mathrm{Var}(\widetilde{\theta})}, \tag{4.15}$$

where $\widetilde{\theta}$ and $\widehat{\theta}$ are the maximum likelihood estimates from (4.9) and (4.10) respectively. For both models, the point estimates from the weighted likelihood (4.10) are slightly different but generally compatible with the ones from the conditional likelihood (4.9). By utilizing the baseline state information and the sampling condition, the standard errors become smaller for estimates of parameters corresponding to the progressive transitions, including NC $\rightarrow$ MCI, MCI $\rightarrow$ dementia, and dementia $\rightarrow$ death. However, estimates of parameters for NC $\rightarrow$ death and MCI $\rightarrow$ NC decreased, and the corresponding standard errors increased very slightly. This may be due to the truncated individuals tending to be "less healthy" than the recruited participants, and therefore more likely to contribute to progressive transitions.

Table 4.2: Results from the Nun Study. Model 1 - a piecewise constant intensity model with a breakpoint at age 90; Model 2 - a piecewise constant intensity model, with a breakpoint at age 90 and risk factor $APOE-\epsilon 4$. RV: relative reduction in variance of the MLE from $\mathcal{L}$ versus $\mathcal{L}_0$. $\alpha_{ij}$: the log ratio of rates for age 90+ for transition from state $i$ to state $j$; $\beta_{ij}$: the log ratio of rates for having $APOE-\epsilon 4$ for transition from state $i$ to state $j$, where $\beta_{i4}, i \in \mathcal{J}$ were fixed at 0.

| | | Conditional, $\mathcal{L}_0$ | | Weighted, $\mathcal{L}$ | | |
|---|---|---|---|---|---|---|
| | Transitions | EST | SE | EST | SE | *RV* % |
| | *Model 1: 90+ v.s. 75 - 90* | | | | | |
| Log Baseline | NC to MCI | -1.290 | 0.073 | -1.132 | 0.057 | 37.215 |
| 75-90 | NC to Death | -3.547 | 0.249 | -3.653 | 0.277 | -9.820 |
| $\log \lambda_{ij0}$ | MCI to NC | -1.866 | 0.082 | -2.086 | 0.086 | -4.235 |
| | MCI to Dementia | -2.627 | 0.104 | -2.692 | 0.089 | 24.692 |
| | MCI to Death | -2.701 | 0.116 | -2.699 | 0.111 | 5.273 |
| | Dementia to Death | -1.669 | 0.098 | -1.647 | 0.087 | 16.934 |
| Log RR, $\geq$90 | NC to MCI | 0.496 | 0.165 | 0.675 | 0.134 | 21.962 |
| $\alpha_{ij}$ | NC to Death | 1.285 | 0.407 | 1.646 | 0.429 | -4.493 |
| | MCI to NC | -0.795 | 0.231 | -0.872 | 0.239 | -2.459 |
| | MCI to Dementia | 0.843 | 0.164 | 1.024 | 0.147 | 17.878 |
| | MCI to Death | 0.817 | 0.174 | 0.788 | 0.173 | 1.794 |
| | Dementia to Death | 0.781 | 0.122 | 0.700 | 0.113 | 11.772 |
| | *Model 2: $APOE-\epsilon 4$ & 90+ v.s. 75 - 90* | | | | | |
| Log Baseline | NC to MCI | -1.302 | 0.079 | -1.143 | 0.064 | 34.924 |
| 75-90, $X=0$ | NC to Death | -3.872 | 0.315 | -3.877 | 0.333 | -11.694 |
| $\log \lambda_{ij0}$ | MCI to NC | -1.706 | 0.088 | -1.875 | 0.090 | -5.370 |
| | MCI to Dementia | -2.782 | 0.122 | -2.941 | 0.106 | 24.051 |
| | MCI to Death | -2.720 | 0.118 | -2.758 | 0.115 | 4.412 |
| | Dementia to Death | -1.727 | 0.104 | -1.690 | 0.093 | 18.652 |
| Log RR, $X=1$ | NC to MCI | 0.227 | 0.191 | 0.152 | 0.142 | 45.155 |
| $\beta_{ij}$ | MCI to NC | -0.944 | 0.260 | -0.858 | 0.262 | -0.917 |
| | MCI to Dementia | 0.531 | 0.179 | 0.615 | 0.136 | 42.104 |
| Log RR, $\geq$90 | NC to MCI | 0.510 | 0.166 | 0.495 | 0.147 | 21.835 |
| $\alpha_{ij}$ | NC to Death | 1.540 | 0.461 | 1.646 | 0.475 | -6.319 |
| | MCI to NC | -0.881 | 0.232 | -0.810 | 0.235 | -2.448 |
| | MCI to Dementia | 0.954 | 0.168 | 1.136 | 0.152 | 18.330 |
| | MCI to Death | 0.799 | 0.179 | 0.821 | 0.178 | 1.275 |
| | Dementia to Death | 0.826 | 0.127 | 0.769 | 0.118 | 12.971 |

## 4.4 Simulation Studies

In this section, we will evaluate the performance of the proposed method through simulation studies. Consider the four-state Markov model as shown in Figure 4.1, and the intensity model

$$\lambda_{ij}(t;X) = \lambda_{ij0}e^{\alpha_{ij}\mathbb{I}\{t>15\}+\beta_{ij}X}, \tag{4.16}$$

and assume that $\beta_{14} = \beta_{24} = \beta_{34} = 0$. Without loss of generality, assume equally spaced yearly assessments over a period of 15 years. The time for entering into the absorbing state is exactly observed. A binary covariate is generated using Bernoulli(0.5). The results are based on 1000 simulations with sample size 300. The left-truncation time (e.g., age at the initial assessment $a_0$) follows Uniform(0, $L$), where $L \in \{5, 15, 25\}$. The baseline transition intensities for $X = 0$ are chosen by controlling the probabilities of disease-free survival at the end of follow-up ($q_1$) and of being in the absorbing state at the end of follow-up time for the oldest group for the case when $x = 0$. ($q_2$).

$$\int \sum_{j=1}^{2}\sum_{k=1}^{2} \pi_j P_{jk}(0,a_0) P_{k1}(a_0, a_0+15)\, da_0 = q_1; \tag{4.17}$$

$$\sum_{j=1}^{3} \pi_j P_{j4}(0, L+15) = q_2. \tag{4.18}$$

The ratios between parameters $\lambda_{120} : \lambda_{140} : \lambda_{210} : \lambda_{230} : \lambda_{240}$ are set as $1 : 1/4 : 1/2 : 2/3 : 3/8$. The initial probability at time 0 is $(0.68, 0.20, 0.12, 0)$.

To see the impact of the range of the left-truncation time, we fix $q_1 = 0.2$ and $q_2 = 0.8$, while assuming the initial probabilities are known. Tables 4.3 gives the results for these simulations, where the biases in the parameter estimates from the conditional likelihood and weighted likelihood approach are negligible. The average of the model based standard errors and empirical standard errors of the estimates from the proposed approach are comparable and consistently smaller than those from the conditional approach, and the shrinkage in the variances becomes larger when the left-truncation range is wider.

Table 4.3: Simulation study comparing performances of the weighted likelihood approach and the conditional approach. Results are based on 1000 replications with sample size 300. CI.cover: coverage probability of 95% confidence intervals.

| | | Conditional Likelihood Approach, $\mathcal{L}_0$ | | | | | Weighted Likelihood Approach, $\mathcal{L}$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TRUE | EST | BIAS% | SE | $SE_{emp}$ | CI.cover | EST | BIAS% | SE | $SE_{emp}$ | CI.cover | $RV$ % |
| *The left-truncation time follows Uniform*$(0, 5)$ | | | | | | | | | | | | |
| $\log \lambda_{120}$ | -3.523 | -3.524 | 0.027 | 0.186 | 0.190 | 0.941 | -3.521 | -0.065 | 0.180 | 0.180 | 0.948 | 6.493 |
| $\log \lambda_{140}$ | -3.928 | -3.884 | -1.136 | 0.158 | 0.149 | 0.938 | -3.883 | -1.145 | 0.157 | 0.149 | 0.932 | 0.411 |
| $\log \lambda_{210}$ | -4.076 | -4.155 | 1.940 | 0.384 | 0.409 | 0.952 | -4.153 | 1.883 | 0.382 | 0.408 | 0.947 | 0.774 |
| $\log \lambda_{230}$ | -3.160 | -3.165 | 0.161 | 0.241 | 0.256 | 0.944 | -3.163 | 0.118 | 0.235 | 0.249 | 0.938 | 4.832 |
| $\log \lambda_{240}$ | -3.671 | -3.642 | -0.764 | 0.255 | 0.252 | 0.945 | -3.642 | -0.780 | 0.254 | 0.251 | 0.946 | 0.872 |
| $\log \lambda_{340}$ | -3.140 | -3.143 | 0.086 | 0.210 | 0.215 | 0.950 | -3.140 | 0.014 | 0.208 | 0.212 | 0.943 | 1.700 |
| $\beta_{12}$ | -0.501 | -0.532 | 6.231 | 0.277 | 0.292 | 0.946 | -0.532 | 6.283 | 0.271 | 0.284 | 0.946 | 4.298 |
| $\beta_{21}$ | 0.877 | 0.919 | 4.835 | 0.447 | 0.467 | 0.951 | 0.921 | 4.968 | 0.444 | 0.465 | 0.952 | 1.063 |
| $\beta_{23}$ | 0.094 | 0.076 | -19.573 | 0.344 | 0.338 | 0.967 | 0.079 | -15.940 | 0.336 | 0.326 | 0.964 | 5.100 |
| $\alpha_{12}$ | 0.588 | 0.547 | -6.959 | 0.362 | 0.350 | 0.966 | 0.545 | -7.350 | 0.360 | 0.348 | 0.966 | 1.068 |
| $\alpha_{14}$ | 0.793 | 0.696 | -12.246 | 0.341 | 0.363 | 0.957 | 0.696 | -12.297 | 0.341 | 0.363 | 0.957 | 0.092 |
| $\alpha_{21}$ | 1.230 | 1.211 | -1.562 | 0.497 | 0.508 | 0.963 | 1.208 | -1.797 | 0.495 | 0.506 | 0.961 | 0.546 |
| $\alpha_{23}$ | 0.824 | 0.770 | -6.584 | 0.424 | 0.439 | 0.966 | 0.769 | -6.770 | 0.421 | 0.440 | 0.963 | 1.069 |
| $\alpha_{24}$ | 0.861 | 0.763 | -11.341 | 0.579 | 0.592 | 0.969 | 0.763 | -11.395 | 0.579 | 0.591 | 0.973 | 0.013 |
| $\alpha_{34}$ | 0.593 | 0.555 | -6.389 | 0.415 | 0.416 | 0.962 | 0.552 | -6.773 | 0.414 | 0.415 | 0.960 | 0.430 |
| *The left-truncation time follows Uniform*$(0, 15)$ | | | | | | | | | | | | |
| $\log \lambda_{120}$ | -3.523 | -3.544 | 0.591 | 0.219 | 0.229 | 0.944 | -3.528 | 0.147 | 0.185 | 0.187 | 0.947 | 28.572 |
| $\log \lambda_{140}$ | -3.928 | -3.886 | -1.087 | 0.197 | 0.198 | 0.936 | -3.885 | -1.093 | 0.194 | 0.196 | 0.937 | 2.820 |
| $\log \lambda_{210}$ | -4.076 | -4.140 | 1.559 | 0.409 | 0.411 | 0.972 | -4.134 | 1.435 | 0.396 | 0.412 | 0.948 | 6.649 |
| $\log \lambda_{230}$ | -3.160 | -3.194 | 1.086 | 0.282 | 0.289 | 0.949 | -3.183 | 0.746 | 0.241 | 0.248 | 0.946 | 26.762 |
| $\log \lambda_{240}$ | -3.671 | -3.669 | -0.048 | 0.324 | 0.338 | 0.941 | -3.654 | -0.461 | 0.313 | 0.324 | 0.946 | 6.730 |
| $\log \lambda_{340}$ | -3.140 | -3.156 | 0.503 | 0.255 | 0.257 | 0.958 | -3.139 | -0.020 | 0.238 | 0.239 | 0.948 | 12.476 |
| $\beta_{12}$ | -0.501 | -0.509 | 1.638 | 0.266 | 0.268 | 0.956 | -0.514 | 2.652 | 0.243 | 0.254 | 0.942 | 16.362 |
| $\beta_{21}$ | 0.877 | 0.890 | 1.504 | 0.395 | 0.396 | 0.962 | 0.900 | 2.653 | 0.387 | 0.407 | 0.940 | 4.452 |
| $\beta_{23}$ | 0.094 | 0.098 | 3.898 | 0.331 | 0.331 | 0.955 | 0.088 | -6.172 | 0.299 | 0.308 | 0.937 | 18.699 |
| $\alpha_{12}$ | 0.588 | 0.595 | 1.139 | 0.269 | 0.277 | 0.950 | 0.587 | -0.258 | 0.249 | 0.256 | 0.949 | 14.559 |
| $\alpha_{14}$ | 0.793 | 0.736 | -7.210 | 0.263 | 0.270 | 0.934 | 0.739 | -6.886 | 0.260 | 0.264 | 0.944 | 1.773 |
| $\alpha_{21}$ | 1.230 | 1.264 | 2.806 | 0.416 | 0.417 | 0.966 | 1.258 | 2.266 | 0.400 | 0.415 | 0.958 | 7.696 |
| $\alpha_{23}$ | 0.824 | 0.839 | 1.703 | 0.334 | 0.344 | 0.948 | 0.844 | 2.354 | 0.304 | 0.310 | 0.953 | 17.408 |
| $\alpha_{24}$ | 0.861 | 0.830 | -3.601 | 0.445 | 0.456 | 0.960 | 0.811 | -5.757 | 0.439 | 0.460 | 0.954 | 2.637 |
| $\alpha_{34}$ | 0.593 | 0.611 | 3.065 | 0.319 | 0.325 | 0.946 | 0.590 | -0.395 | 0.307 | 0.297 | 0.964 | 7.648 |
| *The left-truncation time follows Uniform*$(0, 25)$ | | | | | | | | | | | | |
| $\log \lambda_{120}$ | -3.523 | -3.540 | 0.475 | 0.250 | 0.248 | 0.953 | -3.536 | 0.362 | 0.203 | 0.208 | 0.948 | 34.622 |
| $\log \lambda_{140}$ | -3.928 | -3.905 | -0.601 | 0.237 | 0.250 | 0.932 | -3.892 | -0.936 | 0.231 | 0.235 | 0.931 | 5.207 |
| $\log \lambda_{210}$ | -4.076 | -4.152 | 1.856 | 0.463 | 0.484 | 0.955 | -4.150 | 1.816 | 0.443 | 0.458 | 0.949 | 8.492 |
| $\log \lambda_{230}$ | -3.160 | -3.226 | 2.103 | 0.329 | 0.353 | 0.943 | -3.192 | 1.023 | 0.267 | 0.279 | 0.959 | 34.611 |
| $\log \lambda_{240}$ | -3.671 | -3.671 | 0.008 | 0.391 | 0.418 | 0.946 | -3.669 | -0.049 | 0.383 | 0.423 | 0.941 | -0.679 |
| $\log \lambda_{340}$ | -3.140 | -3.175 | 1.134 | 0.311 | 0.343 | 0.925 | -3.160 | 0.643 | 0.279 | 0.290 | 0.955 | 19.291 |
| $\beta_{12}$ | -0.501 | -0.517 | 3.209 | 0.264 | 0.265 | 0.949 | -0.510 | 1.819 | 0.232 | 0.230 | 0.956 | 22.710 |
| $\beta_{21}$ | 0.877 | 0.896 | 2.139 | 0.389 | 0.398 | 0.942 | 0.894 | 1.908 | 0.374 | 0.395 | 0.942 | 7.341 |
| $\beta_{23}$ | 0.094 | 0.112 | 18.891 | 0.335 | 0.342 | 0.956 | 0.097 | 2.685 | 0.283 | 0.284 | 0.946 | 28.998 |
| $\alpha_{12}$ | 0.588 | 0.599 | 1.856 | 0.281 | 0.278 | 0.961 | 0.593 | 0.846 | 0.251 | 0.252 | 0.957 | 20.595 |
| $\alpha_{14}$ | 0.793 | 0.757 | -4.590 | 0.282 | 0.293 | 0.945 | 0.743 | -6.298 | 0.277 | 0.274 | 0.946 | 3.383 |
| $\alpha_{21}$ | 1.230 | 1.269 | 3.199 | 0.459 | 0.482 | 0.963 | 1.262 | 2.619 | 0.438 | 0.456 | 0.958 | 8.813 |
| $\alpha_{23}$ | 0.824 | 0.868 | 5.288 | 0.360 | 0.378 | 0.953 | 0.848 | 2.817 | 0.316 | 0.316 | 0.955 | 23.510 |
| $\alpha_{24}$ | 0.861 | 0.851 | -1.102 | 0.477 | 0.493 | 0.963 | 0.831 | -3.521 | 0.474 | 0.514 | 0.959 | -1.435 |
| $\alpha_{34}$ | 0.593 | 0.632 | 6.685 | 0.349 | 0.375 | 0.946 | 0.612 | 3.275 | 0.325 | 0.329 | 0.960 | 13.622 |

## 4.5 A Bayesian Approach for Parameter Estimation

When illustrating the method in a frequentist way, we use population prevalences in various cognitive stages as the initial probabilities. Here we consider an alternative way where we acknowledge the uncertainty in its applicability by specifying a prior distribution for the initial probabilities, $\pi_j$'s, and the parameters in the transition intensity model, and use a Bayesian method for parameter estimation.

Table 4.4 gives Bayesian analysis results for the Nun Study. The conditional approach assumes piecewise constant intensities with breakpoint at age 90. The covariate is $APOE - \epsilon 4$. The number of iterations is 2000 (1000 burning + 1000 saved). The prior distributions for the log transition intensities are Normal$(0, 10000)$. The prior distribution for the initial probability is Dirichlet$(c\mu_\pi)$, where the mean $\mu_\pi$ is the population prevalence $(0.68, 0.2, 0.12)$, and the variances are $\mu_\pi(1-\mu_\pi)/(1+c)$ with $c = 5$. Diffuse priors, i.e., Normal$(0, 10000)$ are used for the log transition intensities and log hazard ratios. These choices of priors are applied to several models using the proposed approach including the single-piece and two-piece baseline models and the models with $APOE - \epsilon 4$ as a covariate, and the results are used to compare with the conditional approach where only diffuse priors of the parameters other than the initial probabilities are used. Table 4.4 gives the mean and standard deviations from the posteriors of the parameters. According to the results, the values of the estimated intensities and hazard ratios are similar to those from the frequentist method. This suggests that the model is not very sensitive to a reasonable amount of uncertainty in the auxiliary prevalence data. However, the efficiency gains for the two-piece models, with or without covariate, are not as obvious as before. This is due to the fact that there is variation introduced into the auxiliary information.

Table 4.4: Means and standard deviations of posteriors for the parameter using Bayesian analyses comparing the intensity models for the Nun Study. The prior distribution for prevalence is a Dirichlet distribution with mean $\mu_\pi = (0.68, 0.20, 0.12)$ and variances $\mu_\pi(1-\mu_\pi)/6$. The breakpoint for the two-piece model is age 90.

| | Conditional | | Baseline | | 1-piece, with X | | 2-piece | | 2-piece, with X | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MEAN | SD | MEAN | SD | MEAN | SD | MEAN | SD | MEAN | SD |
| | | | | | *Baseline Transition Intensities* | | | | | |
| $\log \lambda_{120}$ | -1.300 | 0.079 | -1.069 | 0.050 | -1.086 | 0.057 | -1.127 | 0.066 | -1.151 | 0.067 |
| $\log \lambda_{140}$ | -3.936 | 0.328 | -3.314 | 0.214 | -3.578 | 0.283 | -3.927 | 0.366 | -3.922 | 0.361 |
| $\log \lambda_{210}$ | -1.707 | 0.088 | -2.195 | 0.076 | -2.076 | 0.083 | -2.018 | 0.086 | -1.869 | 0.092 |
| $\log \lambda_{230}$ | -2.781 | 0.123 | -2.354 | 0.067 | -2.491 | 0.075 | -2.725 | 0.080 | -2.900 | 0.095 |
| $\log \lambda_{240}$ | -2.734 | 0.131 | -2.430 | 0.086 | -2.435 | 0.082 | -2.749 | 0.118 | -2.752 | 0.115 |
| $\log \lambda_{340}$ | -1.742 | 0.115 | -1.298 | 0.058 | -1.309 | 0.058 | -1.793 | 0.100 | -1.770 | 0.099 |
| | | | | | *APOE–$\epsilon$4* | | | | | |
| $\beta_{12}$ | 0.186 | 0.198 | . | . | 0.161 | 0.140 | . | . | 0.137 | 0.147 |
| $\beta_{21}$ | -0.937 | 0.260 | . | . | -0.791 | 0.256 | . | . | -0.888 | 0.266 |
| $\beta_{23}$ | 0.524 | 0.185 | . | . | 0.479 | 0.129 | . | . | 0.605 | 0.128 |
| | | | | | *Age Group 90+ v.s. 75 - 90* | | | | | |
| $\alpha_{12}$ | 0.511 | 0.166 | . | . | . | . | 0.481 | 0.161 | 0.498 | 0.143 |
| $\alpha_{14}$ | 1.539 | 0.456 | . | . | . | . | 1.623 | 0.525 | 1.591 | 0.504 |
| $\alpha_{21}$ | -0.881 | 0.232 | . | . | . | . | -0.760 | 0.238 | -0.836 | 0.250 |
| $\alpha_{23}$ | 0.951 | 0.169 | . | . | . | . | 0.995 | 0.144 | 1.087 | 0.140 |
| $\alpha_{24}$ | 0.811 | 0.183 | . | . | . | . | 0.812 | 0.182 | 0.811 | 0.185 |
| $\alpha_{34}$ | 0.831 | 0.128 | . | . | . | . | 0.859 | 0.120 | 0.848 | 0.121 |
| | | | | | *Initial Probabilities* | | | | | |
| $\pi_1$ | . | . | 0.773 | 0.075 | 0.732 | 0.089 | 0.635 | 0.090 | 0.635 | 0.087 |
| $\pi_2$ | . | . | 0.213 | 0.077 | 0.251 | 0.091 | 0.352 | 0.092 | 0.355 | 0.088 |
| $\pi_3$ | . | . | 0.014 | 0.013 | 0.017 | 0.014 | 0.013 | 0.012 | 0.010 | 0.012 |

To investigate the robustness of the Bayesian estimates towards the choice of the population prevalence and the variance of the prior for the prevalence, we fit the model using three possible values of the auxiliary prevalence as the mean of the prior distributions for $\pi$. The choices include the population prevalence from the literature; an arbitrary choice of value $(1/3, 1/3, 1/3)$; as well as the equilibrium probabilities of a modified model where every subject entering into the absorbing states will immediately go back to the normal cognition state. Such an equilibrium for the four-state model as shown in figure 4.1 is calculated as the following:

$$\begin{aligned}
\phi_1 &= \frac{\lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}} \\
\phi_2 &= \frac{\lambda_{12}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}} \\
\phi_3 &= \frac{\lambda_{12}\lambda_{23}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}}.
\end{aligned} \tag{4.19}$$

Details of the derivation of these results are given in Appendix B.3.

Recall the variance of the prior of $\pi$ is $\mu_\pi(1-\mu_\pi)/(1+c)$ . When increasing the value of $c$, the variance becomes smaller. Table 4.5 gives the results where $c$ is valued at 1, 3 and 5. Based on the results, the means and standard deviations from the posteriors are quite stable with respect to changing the amount of variation in prior distributions of $\pi$.

The means of the posteriors from the Bayesian analyses are very close to the frequentist point estimates. However, most of the posterior standard deviations are slightly larger than the frequentist standard errors. This may be due to that the Bayesian approach is treating the initial probabilities as unknown, as well as the uncertainty in the prior information. However, since the Bayesian analysis is pretty robust towards the prior information on the prevalence data, it has the advantage of accommodating misspecification and uncertainty in the auxiliary prevalence data and allows for direct inference from the posterior.

Table 4.5: Means and standard deviations of posteriors for the parameter using Bayesian analyses comparing the Dirichlet priors for the prevalence for the Nun Study. The mean of the prior distribution, $\mu_\pi$, is one of the three scenarios: Population prevalence - $(0.68, 0.20, 0.12)$; Reborn equilibrium probabilities - $(0.44, 0.42, 0.14)$; Evenly distributed - $(0.33, 0.33, 0.33)$. The variances of the prior distribution are $\mu_{\pi,k}(1-\mu_{\pi,k})/(1+c)$.

| | Conditional | Population Prevalence | | Equilibrium Probabilities | | Evenly Distributed | |
|---|---|---|---|---|---|---|---|
| | MEAN | MEAN | SD | MEAN | SD | MEAN | SD |
| | | | | *Variance Coefficient:* $c = 1$ | | | |
| $\log \lambda_{120}$ | -1.213 | -1.075 | 0.050 | -1.077 | 0.051 | -1.073 | 0.059 |
| $\log \lambda_{140}$ | -3.345 | -3.297 | 0.210 | -3.293 | 0.210 | -3.305 | 0.227 |
| $\log \lambda_{210}$ | -2.044 | -2.196 | 0.076 | -2.199 | 0.079 | -2.198 | 0.079 |
| $\log \lambda_{230}$ | -2.315 | -2.346 | 0.064 | -2.349 | 0.071 | -2.356 | 0.065 |
| $\log \lambda_{240}$ | -2.417 | -2.420 | 0.077 | -2.428 | 0.089 | -2.422 | 0.081 |
| $\log \lambda_{340}$ | -1.241 | -1.300 | 0.054 | -1.304 | 0.059 | -1.302 | 0.055 |
| $\pi_1$ | | 0.750 | 0.093 | 0.758 | 0.084 | 0.760 | 0.087 |
| $\pi_2$ | | 0.250 | 0.093 | 0.239 | 0.084 | 0.230 | 0.089 |
| $\pi_3$ | | 0.000 | 0.000 | 0.003 | 0.006 | 0.011 | 0.012 |
| | | | | *Variance Coefficient:* $c = 3$ | | | |
| $\log \lambda_{120}$ | -1.213 | -1.077 | 0.050 | -1.078 | 0.052 | -1.075 | 0.052 |
| $\log \lambda_{140}$ | -3.345 | -3.300 | 0.207 | -3.308 | 0.223 | -3.307 | 0.222 |
| $\log \lambda_{210}$ | -2.044 | -2.198 | 0.077 | -2.198 | 0.077 | -2.201 | 0.078 |
| $\log \lambda_{230}$ | -2.315 | -2.356 | 0.067 | -2.358 | 0.066 | -2.362 | 0.065 |
| $\log \lambda_{240}$ | -2.417 | -2.428 | 0.082 | -2.420 | 0.080 | -2.419 | 0.086 |
| $\log \lambda_{340}$ | -1.241 | -1.304 | 0.057 | -1.304 | 0.057 | -1.301 | 0.060 |
| $\pi_1$ | | 0.754 | 0.084 | 0.729 | 0.079 | 0.740 | 0.092 |
| $\pi_2$ | | 0.237 | 0.088 | 0.262 | 0.083 | 0.238 | 0.094 |
| $\pi_3$ | | 0.009 | 0.012 | 0.010 | 0.017 | 0.021 | 0.014 |
| | | | | *Variance Coefficient:* $c = 5$ | | | |
| $\log \lambda_{120}$ | -1.213 | -1.068 | 0.050 | -1.079 | 0.052 | -1.077 | 0.051 |
| $\log \lambda_{140}$ | -3.345 | -3.292 | 0.203 | -3.284 | 0.208 | -3.296 | 0.210 |
| $\log \lambda_{210}$ | -2.044 | -2.192 | 0.076 | -2.199 | 0.077 | -2.195 | 0.078 |
| $\log \lambda_{230}$ | -2.315 | -2.352 | 0.067 | -2.358 | 0.065 | -2.361 | 0.064 |
| $\log \lambda_{240}$ | -2.417 | -2.427 | 0.085 | -2.426 | 0.087 | -2.420 | 0.085 |
| $\log \lambda_{340}$ | -1.241 | -1.296 | 0.058 | -1.295 | 0.056 | -1.295 | 0.055 |
| $\pi_1$ | | 0.773 | 0.075 | 0.730 | 0.079 | 0.732 | 0.071 |
| $\pi_2$ | | 0.213 | 0.077 | 0.251 | 0.079 | 0.244 | 0.071 |
| $\pi_3$ | | 0.014 | 0.013 | 0.018 | 0.019 | 0.024 | 0.016 |

## 4.6 Discussion

In this chapter, we proposed a weighted likelihood approach for disease process data while utilizing the auxiliary information on disease prevalence data. The proposed method is able to adjust for the left-truncation problem in multistate models. We described the method details for time-homogeneous and piecewise constant Markov models under intermittent observation.

We have focused on modeling the transition intensities and risk factor effects for dementia using the Nun Study data. The proposed method successfully improved efficiency for parameters related to progressive transitions. Simulation studies suggested that the efficiency gain will increase when the left-truncation problem is stronger. However, incorporating auxiliary prevalence cannot ensure efficiency gain for all of the parameters. Which parameters are impacted in the opposite way depends on the structure of the disease process model.

To account for uncertainty in the prevalence data, we have also used a Bayesian method to get estimates. Bayesian analyses give similar point estimates with slightly less efficiency gain. When prevalence data are not directly available, we proposed an equilibrium probability for the reborn process to simulate a snapshot of a population with incoming and outgoing individuals. This is done by constructing a reborn process corresponding to the disease process with absorbing state(s) using certain "reborn" rules.

In conclusion, the proposed weighted likelihood method can adjust for the left-truncation problem in multistate models which is a common issue in disease process data due to the fact that in most cases,, the individuals have to be alive to enroll in the study. However, valid estimates will depend on reliable auxiliary information. One may want to use Bayesian analyses in case of uncertainty in the auxiliary prevalence data. There are a number of interesting future research directions for this topic. We will discuss them in details in Chapter 5.

# Chapter 5

# Discussion and Future Work

In this thesis, we developed novel statistical methods for time-to-event analysis and multi-state models to deal with left-truncation and response-dependent sampling problems using auxiliary information.

In chapter 2, we incorporated auxiliary information from census and external studies on the same cohort to a sample with large portion of left-truncation. We derived the observed likelihood for both the sample and auxiliary information. In case where the closed form of the likelihood is intractable, an Monte-Carlo approach was proposed to approximate the values of the score and information function. Illustrated by both simulation and real-data example, the proposed method improves the estimation efficiency of the survival function comparing to the conditional likelihood approach without use of auxiliary information.

In chapter 3, we developed a finite mixture hidden Markov model to relax the Markov assumption and consider multiple disease types with common symptomatic stages. The proposed method is also able to incorporate auxiliary information on disease typesfor some study subjects. . Through real-data analysis and simulation , we found that the proposed method is able to identify different types of the underlying disease and gives type-specific and overall prevalence estimates using either a frequentist or a Bayesian approach. We also investigated the potential estimability and identifiability issues associated with the proposed method and explored different ways to deal with them.

In chapter 4, we addressed the state-dependent sampling problem in muitistate models analyzing disease process data. We proposed a weighted approach which considers the sampling condition and makes use of population age-specific disease prevalence data. The maximum likelihood estimates were obtained by used both frequentist and Bayesian approaches. We found that incorporating baseline status information given the sampling condition improves the estimation efficiency for most of the unknown parameters.

In the following sections, we provide some general discussion and propose future work for each chapter.

## 5.1 Future Work on Augmented Likelihood for Incorporating Auxiliary Information

In chapter 2, we focused on parametric models for the event times. However non-parametric estimates can be considered using a similar approach. Extension to a nonparametric model can be done by assigning nonparametric forms of the $g$ functions and the conditional likelihood in (2) and (4). This can be more straightforward when combining with the one-step MCEM algorithm which doesn't require the derivation of the $g$ functions. If the conditional estimates doesn't cover the range of the pseudo-data, values of the pseudo-data can be generated subject to only the auxiliary yearly number of deaths.

Another extension is to model the effects of risk factors. Incorporating covariates into the method can be challenging if we require the pseudo-values to have exactly the same components as the observed data. However, if the covariate effects for the truncated part are not of interest, which is reasonable to assume, we can focus on generating the baseline survival time for the pseudo-data and ignore the covariate effects in the form of the likelihood functions for the pseudo-data set $\mathcal{D}_p$.

If we consider the combination of data from the truncated sample and the observed sample, $(\mathcal{D}_p, \mathcal{D})$, as the complete data, and the truncated sample $\mathcal{D}_p$ as the missing data, the procedure using the Monte-Carlo Expectation to approximate the score function can be viewed as a special case of multiple imputation. Therefore, inference schemes for multiple

imputation can be applied to the proposed method, such as the Rubin's variance estimator (Rubin, 2004), which will take the extra variation caused by the generation of the truncated sample into consideration. Letting $\text{Var}(\hat{\theta}^{(r)})$ be the variance estimate for the MLE from the $r^{\text{th}}$ combined sample $(\mathcal{D}_p, \mathcal{D})$, the Rubin's variance estimator is

$$\begin{aligned} \hat{\Sigma}_{\text{Rubin}} &= \frac{1}{R}\sum_{r=1}^{R}\text{Var}(\hat{\theta}^{(r)}) + \frac{R+1}{R}\frac{1}{R-1}\sum_{r=1}^{R}(\hat{\theta}^{(r)} - \hat{\theta})^2 \\ &= W + (1+R^{-1})B \end{aligned} \tag{5.1}$$

where $W$ is the within-imputation variance and $B$ the between-imputation variance. The number of degrees of freedom for the Student's t-distribution is

$$df = (R-1)\left(1 + \frac{W}{(1+R^{-1})B}\right)^2. \tag{5.2}$$

With big enough $R$, the confidence interval can be approximated using a normal distribution. However, when applying 5.1 to the proposed likelihood, we found that it tends to overestimate the variance. This may be dues to that the pseudo-values are generated using a frequency-based method. Therefore, we consider to test a robust alternative to (5.1) proposed by Rubin (2004).

There are other applications of the proposed method. One example is the Honolulu-Asia Aging Study (Higuchi et al., 2015; Huh et al., 2015), which began in 1991. The cohort, with 3734 participants, is a subsample of the cohort in the Honolulu Heart Program, which began in 1965. Assuming the research interest lies on modeling the survival distribution or the time-to-onset of a disease, the summary level information for the Honolulu Heart Program can be incorporated in the estimation procedure.

In addition to time-to-event data, other event history data are also subject to the left-truncation problem. For example, auxiliary information regarding the truncated population may be obtained from the same or similar cohort for multistate models. One can derive a similar likelihood to (2.4) by determining the form of the likelihood contributions according to the model and the forms of the auxiliary information. Again, the value of

the likelihood can be approximated using the Monte-Carlo expectations as long as one can generate pseudo-values to reflect the auxiliary information.

## 5.2 Future Work on Finite Mixture Hidden Markov Models

An ongoing work for chapter 3 is adding risk factors to the transition intensities. Let $X = (X_1, X_2, \ldots, X_P)$ be the vector of the risk factors. For underlying disease type $m = 1, 2$, let $\beta_{ij}^{(m)} = (\beta_{ij,1}^{(m)}, \beta_{ij,2}^{(m)}, \ldots, \beta_{ij,P}^{(m)})$ be the log instantaneous rate ratios for transition from state $i$ to state $j$. The proportional hazard model for transition intensities of disease type $m$ is

$$\lambda_{ij}^{(m)}(t; \mathrm{X}) \quad = \quad \lambda_{ij0}^{(m)}(t) e^{X^{\mathrm{T}} \beta_{ij}^{(m)}}. \tag{5.3}$$

If the difference between covariate effects between disease types is not of interest, one can assume $\beta_{ij}^{(1)} = \beta_{ij}^{(2)} = \beta_{ij}$. We found that putting constraints on parameters or the ratios between the parameters may solve the identifiability and estimability issues. In multistate models, if the states are defined as the disease stages or have other intuitive interpretations, auxiliary information on these parameters or ratios may be obtained from the literature. For instance, there may be information on the relative risk of mortality between patients with different types of dementia, which corresponds to $\lambda_{23}^{(1)}/\lambda_{45}^{(2)}$. This information can be definite numbers, approximate intervals or prior distributions.

Another extension is to model the missingness mechanism of the mixture component indicator. In the Nun Study, the selection criteria for conducting autopsy is assumed to be random, which makes it reasonble to assume that the missing in the pathology data is completely at random. However, in the case where the missingness depends on the disease stages, jointly modeling the transition intensities and the mixture components is required.

The proposed method can be extended to cases where the number of components in the mixture is larger than two. However, a more complex model requires more attention

to the potential identifiability and estimability issues. For the estimability issue, alternative algorithms or changing initial values may help with the problem. In terms of the identifiability, putting constraints on parameters or simplifying the model will be a good consideration to solve the issue. If one's interest is primarily in predicting the observed process, rather than interpreting the estimates, a model with an identifiability problem may still give valid prediction of the event time.

## 5.3 Future Work on Left-truncated Multistate Models

In chapter 4, we have considered the auxiliary information in form of a constant value or a prior distribution that is identical for the entire sample. Sometimes the prevalence may depend on covariates in the dataset. In this case, pre-built models or prior information from literature may be borrowed. For example Brookmeyer et al. (2018) proposed a multistate model to forecast the yearly prevalence for MCI, preclinical and clinical Alzheimer's disease, from considering the impacts of AD pathology. Barnes and Yaffe (2011) proposed a model using the relative risks of Alzheimer's disease versus other modifiable risk factors including cardiovascular risk factors, psychosocial factors, and health behaviours, such as diabetes, depression, and smoking. Norton et al. (2014) provided specific estimates of preventive potential by accounting for the association between risk factors and discussed the concerns raised by Barnes and Yaffe (2011). In addition, the propose weighted likelihood is conditioning on left-truncation time. Therefore, another possible extension is to utilize auxiliary information on the left-truncation time distribution for the sample, i.e. $f(a_0 \mid Z(a_0) \neq 4)$. In this case, the weight in (4.11) will take a form of a joint probability for $a_0$ and $Z(a_0)$ conditioning on being alive at accrual.

Another extension is relaxing the Markov assumption on the multistate model. For instance, in the proposed mixture Hidden Markov model in chapter 3, we estimated the initial probabilities for being in a disease state at accrual and the transition intensities simultaneously. Assume that the auxiliary prevalence data are available for the age at

accrual. Letting $\mathcal{J}_D$ be the set of absorbing states, formula (3.10) can be rewritten as

$$P\left(Y\left(t_0\right)=y_0 \mid Z\left(t_0\right)=j\right) P\left(Z\left(t_0\right)=j \mid \text{alive at } t_0\right) \quad = \quad e_{jy_0} w_Z\left(t_0, j\right), \tag{5.4}$$

where

$$w_Z\left(t_0, j\right) \quad = \quad \frac{\sum_{u \notin \mathcal{J}_D} P\left(Z(t_0) | Z(0)=u, t_0\right) \pi_u}{\sum_{u \notin \mathcal{J}_D} P\left(Z(t_0) \notin \mathcal{J}_D | Z(0)=u, t_0\right) \pi_u}. \tag{5.5}$$

Thus formulas (3.10 - 3.14) will contain only the mixture probability and the transition intensities as unknown parameter. This may improve the estimate effciency and help with the identifiability and estimability issues.

Other extensions include the semi-Markov model and the time-to-event models, where the sampling conditions are usually in terms of truncation or prevalence of states.

# Bibliography

P. K. Andersen. Multistate models in survival analysis: A study of nephropathy and mortality in diabetes. *Statistics in Medicine*, 7(6):661–670, 1988.

P. K. Andersen and N. Keiding. Multi-state models for event history analysis. *Statistical Methods in Medical Research*, 11(2):91–115, 2002.

M. Asgharian, C. E. M'Lan, and D. B. Wolfson. Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association*, 97 (457):201–209, 2002.

D. E. Barnes and K. Yaffe. The projected effect of risk factor reduction on Alzheimer's disease prevalence. *The Lancet Neurology*, 10(9):819–828, 2011.

M. Berzofsky and P. P. Biemer. Weak identifiability in latent class analysis. In *Proceedings of the ASA Section on Survey Methodology, Joint Statistical Meetings*, pages 4346–4354, 2012.

R. Brookmeyer, N. Abdalla, C. H. Kawas, and M. M. Corrada. Forecasting the prevalence of preclinical and clinical Alzheimer's disease in the united states. *Alzheimer's & Dementia*, 14(2):121–129, 2018.

S. M. Butler and D. A. Snowdon. Trends in mortality in older women: Findings from the Nun study. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 51(4):S201–S208, 1996.

Census Bureau of the United States. United States National Intercensal Tables: 1900-1990, 1900-1990. URL https://www.census.gov/data/tables/time-series/demo/popest/pre-1980-national.html.

Centers for Disease Control and Prevention. National Vital Statistics Reports, Aug 2017. URL https://www.cdc.gov/nchs/products/nvsr.htm.

D. Commenges. Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research*, 11(2):167–182, 2002.

D. Commenges, P. Joly, L. Letenneur, and J.-F. Dartigues. Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in Medicine*, 23 (2):199–210, 2004.

A. J. Copas and V. T. Farewell. Incorporating retrospective data into an analysis of time to illness. *Biostatistics*, 2(1):1–12, 2001.

C. L. Faucett, N. Schenker, and J. M. Taylor. Survival analysis using auxiliary variables via multiple imputation, with application to AIDS clinical trial data. *Biometrics*, 58(1): 37–47, 2002.

Y. Foucher, E. Mathieu, P. Saint-Pierre, J.-F. Durand, and J.-P. Daurès. A semi-Markov model based on generalized Weibull distribution with an illustration for hiv disease. *Biometrical Journal*, 47(6):825–833, 2005.

Y. Foucher, M. Giral, J.-P. Soulillou, and J.-P. Daures. A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Statistics in Medicine*, 26(30):5381–5393, 2007.

H. Frydman. Nonparametric estimation of a Markov illness-death process from interval-censored observations, with application to diabetes survival data. *Biometrika*, 82(4): 773–789, 1995.

R. Gentleman, J. Lawless, J. Lindsey, and P. Yan. Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine*, 13(8):805–821, 1994.

M. Higuchi, R. Chen, R. D. Abbott, C. Bell, L. Launer, G. W. Ross, H. Petrovitch, and K. Masaki. Mid-life proteinuria and late-life cognitive function and dementia in elderly men: The Honolulu-Asia Aging Study. *Alzheimer Disease and Associated Disorders*, 29 (3):200, 2015.

M. G. Hudgens. On nonparametric maximum likelihood estimation with interval censoring and left truncation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(4):573–587, 2005.

J. Y. Huh, G. W. Ross, R. Chen, R. D. Abbott, C. Bell, B. Willcox, L. Launer, H. Petrovitch, B. Kaya, and K. Masaki. Total and differential white blood cell counts in late life predict 8-year incident stroke: The Honolulu Heart Program. *Journal of the American Geriatrics Society*, 63(3):439–446, 2015.

C. H. Jackson et al. Multi-state models for panel data: The msm package for R. *Journal of Statistical Software*, 38(8):1–29, 2011.

S. L. Jiwani. Parametric changepoint survival model with application to coronary artery bypass graft surgery data. Master's thesis, Department of Statistics and Actuarial Science, Simon Fraser University, 2005.

P. Joly and D. Commenges. A penalized likelihood approach for a progressive three-state model with censored and truncated data: Application to AIDS. *Biometrics*, 55(3):887–890, 1999.

J. Kalbfleisch and J. Lawless. Regression models for right truncated data with applications to AIDS incubation times and reporting lags. *Statistica Sinica*, 1(1):19–32, 1991.

J. Kalbfleisch and J. F. Lawless. The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80(392):863–871, 1985.

M. Kang and S. W. Lagakos. Statistical methods for panel data from a semi-Markov process, with application to HPV. *Biostatistics*, 8(2):252–264, 2006.

N. Keiding and M. Moeschberger. *Independent Delayed Entry, Survival Analysis: State of the Art*, volume 211. Springer, 1992.

G. Li and J. Qin. Semiparametric likelihood-based inference for biased and truncated data when the total sample size is known. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(1):243–254, 1998.

T. A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 44(2):226–233, 1982.

S. Norton, F. E. Matthews, D. E. Barnes, K. Yaffe, and C. Brayne. Potential for primary prevention of alzheimer's disease: an analysis of population-based data. *The Lancet Neurology*, 13(8):788–794, 2014.

J. Qin and Y. Shen. Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics*, 66(2):382–392, 2010.

J. Qin, J. Ning, H. Liu, and Y. Shen. Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association*, 106(496):1434–1449, 2011.

K. P. Riley, D. A. Snowdon, and W. R. Markesbery. Alzheimer's neurofibrillary pathology and the spectrum of cognitive function: findings from the Nun Study. *Annals of neurology*, 51(5):567–577, 2002.

D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys*, volume 81. John Wiley & Sons, 2004.

G. A. Satten and M. R. Sternberg. Fitting semi-Markov models to interval-censored data with unknown initiation times. *Biometrics*, 55(2):507–513, 1999.

P.-S. Shen. Nonparametric estimation with left-truncated and right-censored data when the sample size before truncation is known. *Statistics*, 48(2):315–326, 2014.

P. Smyth. Hidden Markov models for fault detection in dynamic systems. *Pattern Recognition*, 27(1):149–164, 1994.

D. A. Snowdon. Healthy aging and dementia: Findings from the Nun Study. *Annals of Internal Medicine*, 139(5 Pt 2):450–454, 2003.

D. A. Snowdon, L. H. Greiner, J. A. Mortimer, K. P. Riley, P. A. Greiner, and W. R. Markesbery. Brain infarction and the clinical expression of Alzheimer disease: The Nun Study. *JAMA*, 277(10):813–817, 1997.

A. C. Titman. Estimating parametric semi-Markov models from panel data using phase-type approximations. *Statistics and Computing*, 24(2):155–164, 2014.

A. C. Titman and L. D. Sharples. Semi-Markov models with phase-type sojourn distributions. *Biometrics*, 66(3):742–752, 2010.

W.-Y. Tsai, N. P. Jewell, and M.-C. Wang. A note on the product-limit estimator under right censoring and left truncation. *Biometrika*, 74(4):883–886, 1987.

N. B. Tuma, M. T. Hannan, and L. P. Groeneveld. Dynamic analysis of event histories. *American Journal of Sociology*, 84(4):820–854, 1979.

B. Vakulenko-Lagun and M. Mandel. Comparing estimation approaches for the illness–death model under left truncation and right censoring. *Statistics in Medicine*, 35(9): 1533–1548, 2016.

Y. Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika*, 76(4):751–761, 1989.

M.-C. Wang. Product limit estimates: A generalized maximum likelihood study. *Communications in Statistics - Theory and Methods*, 16(11):3117–3132, 1987.

M.-C. Wang. Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association*, 86(413):130–143, 1991.

M.-C. Wang. Hazards regression analysis for length-biased data. *Biometrika*, 83(2):343–354, 1996.

M.-C. Wang, N. P. Jewell, and W.-Y. Tsai. Asymptotic properties of the product limit estimate under random truncation. *Annals of Statistics*, 14(4):1597–1605, 1986.

M.-C. Wang, R. Brookmeyer, and N. P. Jewell. Statistical models for prevalent cohort data. *Biometrics*, 49(1):1–11, 1993.

J. A. Yesavage, R. O'Hara, H. Kraemer, A. Noda, J. L. Taylor, S. Ferris, M.-C. Gely-Nargeot, A. Rosen, L. Friedman, J. Sheikh, et al. Modeling the prevalence and incidence of Alzheimer's disease and mild cognitive impairment. *Journal of Psychiatric Research*, 36(5):281–286, 2002.

# Appendix A

# Appendix for Chapter 3

## A.1 Transition Probabilities for the Underlying Models

### A.1.1 Closed Forms for Disease Type One

$$\begin{aligned} p_{11}(s,t) &= e^{-h_1(t-s)} \\ p_{12}(s,t) &= \begin{cases} -\frac{\lambda_{12}}{h_1-\lambda_{23}}\left[e^{-h_1(t-s)} - e^{-\lambda_{23}(t-s)}\right] & \text{if } h_1 \neq \lambda_{23} \\ \lambda_{12}(t-s)\, e^{-\lambda_{23}(t-s)} & \text{if } h_1 = \lambda_{23} \end{cases} \\ p_{22}(s,t) &= e^{-\lambda_{23}(t-s)} \end{aligned}$$

$p_{21}(s,t) = p_{31}(s,t) = p_{32}(s,t) = 0$, and $p_{i3}(s,t) = 1 - \sum_{j=1}^{2} p_{ij}(s,t)$.

### A.1.2 Closed Forms for Disease Type Two

$$
\begin{aligned}
p_{11}(s,t) &= e^{-h_1(t-s)} \\
p_{12}(s,t) &= \begin{cases} -\frac{\lambda_{12}}{h_1-h_2}\left[e^{-h_1(t-s)}-e^{-h_2(t-s)}\right], & \text{if } h_1 \neq h_2 \\ \lambda_{12}(t-s)\,e^{-h_1(t-s)}, & \text{if } h_1 = h_2 \end{cases} \\
p_{13}(s,t) &= \begin{cases} \frac{\lambda_{12}\lambda_{23}e^{-h_1(t-s)}}{(h_1-h_2)(h_1-\lambda_{35})} - \frac{\lambda_{12}\lambda_{23}e^{-h_2(t-s)}}{(h_1-h_2)(h_2-\lambda_{35})} + \frac{\lambda_{12}\lambda_{23}e^{-\lambda_{35}(t-s)}}{(h_1-\lambda_{35})(h_2-\lambda_{35})}, & \text{if } h_1 \neq h_2 \neq \lambda_{35} \\ -\frac{\lambda_{12}\lambda_{23}}{(h_1-\lambda_{35})^2}\left[e^{-h_1(t-s)}-e^{-\lambda_{35}(t-s)}\right] - \frac{\lambda_{12}\lambda_{23}}{(h_1-\lambda_{35})}(t-s)\,e^{-h_1(t-s)}, & \text{if } h_1 = h_2 \neq \lambda_{35} \\ \frac{\lambda_{12}\lambda_{23}}{(h_2-\lambda_{35})^2}\left[e^{-h_2(t-s)}-e^{-\lambda_{35}(t-s)}\right] + \frac{\lambda_{12}\lambda_{23}}{h_2-\lambda_{35}}(t-s)\,e^{-\lambda_{35}(t-s)}, & \text{if } h_2 \neq h_1 = \lambda_{35} \\ \frac{\lambda_{12}\lambda_{23}}{(h_1-\lambda_{35})^2}\left[e^{-h_1(t-s)}-e^{-\lambda_{35}(t-s)}\right] + \frac{\lambda_{12}\lambda_{23}}{h_1-\lambda_{35}}(t-s)\,e^{-\lambda_{35}(t-s)}, & \text{if } h_1 \neq h_2 = \lambda_{35} \\ \frac{\lambda_{12}\lambda_{23}}{2}(t-s)^2\,e^{-\lambda_{35}(t-s)} & \text{if } h_1 = h_2 = \lambda_{35} \end{cases} \\
p_{14}(s,t) &= \frac{\lambda_{14}}{h_1}\left(1-e^{-h_1(t-s)}\right) \\
p_{22}(s,t) &= e^{-h_2(t-s)} \\
p_{23}(s,t) &= \begin{cases} -\frac{\lambda_{23}}{h_2-\lambda_{35}}\left[e^{-h_2(t-s)}-e^{-\lambda_{35}(t-s)}\right], & \text{if } h_2 \neq \lambda_{35} \\ \lambda_{23}(t-s)\,e^{-\lambda_{35}(t-s)}, & \text{if } h_2 = \lambda_{35} \end{cases} \\
p_{33}(s,t) &= e^{-\lambda_{35}(t-s)}
\end{aligned}
$$

$p_{21}(s,t) = p_{24}(s,t) = p_{31}(s,t) = p_{32}(s,t) = p_{34}(s,t) = p_{41}(s,t) = p_{42}(s,t) = p_{43}(s,t) = p_{45}(s,t) = p_{51}(s,t) = p_{52}(s,t) = p_{53}(s,t) = p_{54}(s,t) = 0$, and $p_{i5}(s,t) = 1 - \sum_{j=1}^{4} p_{ij}(s,t)$.

## A.2 Prevalence of Dementia for the Nun Study

Let $f(t)$ be the prevalence for dementia at time $t$ for individuals shown in Figure 3.2, which is the proportion of individuals who have experienced the dementia state by time $t$. Then

$$f(t) = P\left(\exists s \leq t, Y(s) = 2\right) = \psi_1 f_1(t) + \psi_1 f_2(t), \tag{A.1}$$

where the prevalence for dementia for disease type 1 and 2 are

$$\begin{aligned} f_1(t) &= P\left(\exists s \leq t, Z^{(1)}(s) = 2\right); \\ f_2(t) &= P\left(\exists s \leq t, Z^{(2)}(s) = 3\right). \end{aligned} \tag{A.2}$$

Calculation of $f_1(t)$ is not straightforward using transition probabilities of the underlying model $Z^{(1)}(t)$. Alternatively, consider a Markov model with distinct death state for each of the transient state, see Figure A.1. The transition intensities of the new process $W^{(1)}(t)$ remain the same as the original process $Z^{(1)}(t)$. It's obvious that $W^{(1)}(t)$ has the same prevalence for dementia as $Z^{(1)}(t)$ at any $t$. The transition intensity matrix of $W^{(1)}(t)$

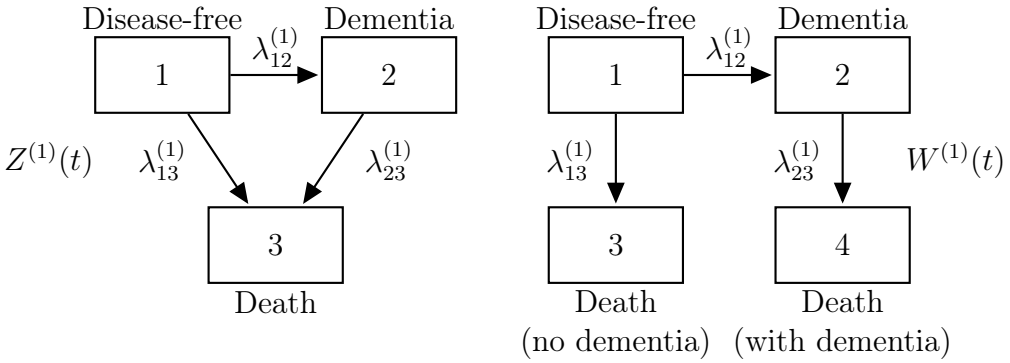


Figure A.1: A transformed multistate process for calculating the cumulative prevalence for dementia - Disease type 1. $Z^{(1)}(t)$: The original process; $W^{(1)}(t)$: A process with distinct absorbing state for each transient state.

is

$$\Lambda_W^{(1)} = \begin{bmatrix} -\lambda_{12}^{(1)} - \lambda_{13}^{(1)} & \lambda_{12}^{(1)} & \lambda_{13}^{(1)} & 0 \\ 0 & 0 & -\lambda_{23}^{(1)} & \lambda_{23}^{(1)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The transition probability matrix for $W^{(1)}(t)$ is $\mathrm{P}_W^{(1)}(s,t) = e^{(t-s)\Lambda_W^{(1)}}$. For such a model, individuals who experienced the dementia state by time $t$ will be in either state 2 or 4 at time $t$. Recall that $\pi_i^{(1)}$ is the probability for being in state $i$ at time 0, $i = 1, 2$. Then

$$f_1(t) = \begin{bmatrix} \pi_1^{(1)} & \pi_2^{(1)} & 0 & 0 \end{bmatrix} \mathrm{P}_W^{(1)}(0,t) \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}. \tag{A.3}$$

Similarly for disease type 2, the new stochastic process $W^{(2)}(t)$ with distinct absorbing states has the form in Figure A.2. where $\mathrm{P}_W^{(2)}(s,t) = e^{(t-s)\Lambda_W^{(2)}}$. The transition intensity

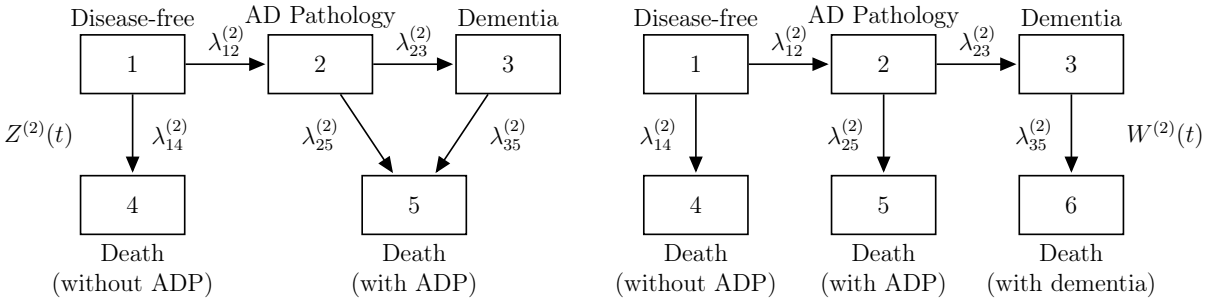


Figure A.2: A transformed multistate process for calculating the cumulative prevalence for dementia - Disease type 2. $Z^{(2)}(t)$: The original process; $W^{(2)}(t)$: A process with distinct absorbing state for each transient state.
*ADP: Alzheimer's disease pathology.

matrix for $W^{(2)}(t)$ is

$$\Lambda_W^{(2)} = \begin{bmatrix} -\lambda_{12}^{(2)} - \lambda_{14}^{(2)} & \lambda_{12}^{(2)} & 0 & \lambda_{14}^{(2)} & 0 & 0 \\ 0 & -\lambda_{23}^{(2)} - \lambda_{25}^{(2)} & \lambda_{23}^{(2)} & 0 & \lambda_{25}^{(2)} & 0 \\ 0 & 0 & 0 & 0 & -\lambda_{36}^{(6)} & \lambda_{36}^{(6)} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

For $W^{(2)}(t)$, individuals who experienced the dementia state by time $t$ will be in either state 3 or 6 at time $t$. Then

$$f_2(t, \pi^{(2)}) = \begin{bmatrix} \pi_1^{(2)} & \pi_2^{(2)} & \pi_3^{(2)} & 0 \end{bmatrix} \mathrm{P}_W^{(2)}(0,t) \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \tag{A.4}$$

The prevalence using (A.1) allows individuals to have dementia at $t = 0$ (i.e., age 75 in the Nun Study). If one is interested in the prevalence for people who are in the disease-free state without AD pathology at age 75, i.e.,

$$\psi_1 P\left(\exists s \le t, Z^{(1)}(s) = 2 \mid Z^{(1)}(0) = 1\right) + \psi_2 P\left(\exists s \le t, Z^{(2)}(s) = 3 \mid Z^{(2)}(0) = 1\right), \tag{A.5}$$

this can be calculated by substituting $\pi^{(1)} = (1, 0, 0)$ and $\pi^{(2)} = (1, 0, 0, 0, 0)$ into (A.1). Curves in Figure 3.4 are using (A.5).

## A.3 The Likelihood Function for the Nun Study

To investigate the identifiability issue and conduct Bayesian analyses using JAGS, we derived the closed forms of the likelihood function for the model in Figure 3.2.

Define the $a_i, i = 1, 2, ..., 5$ as shown in Table A.1, where each $a_i$ values at one the assessment times $t_0, t_1, ..., t_K$ or the time at death $T_D$.

The form of the likelihood function for one participant is determined by the observed transition path and the autopsy result for AD pathology of this participant. Table A.2 gives a list of possible combinations of the transitions paths and pathology.

Consider $\mathcal{L} = \mathcal{L}^{(1)} + \mathcal{L}^{(2)}$, where $\mathcal{L}^{(m)}$ represent the contributions from underlying disease type $m$. Table A.3 gives the case-wise likelihood in terms of the productions of transition probabilities. The closeforms are given in the following pages.

Table A.1: Selected assessment times essential for derivation of the likelihood function

| Notation | Definition |
|---|---|
| $a_1$ | Time at the first assessement in the disease-free state |
| $a_2$ | Time of last assessement in the disease-free state |
| $a_3$ | Time at the first assessement in the dementia state |
| $a_4$ | Time of last assessement in the dementia state |
| $a_5$ | Time at death, i.e. $a_5 = T_D$ |
| $a_{ij}$ | $a_j - a_i$, for $i, j = 1, 2, ..., 5$ and $i < j$ |

Table A.2: Cases of participants by observed transition paths and the autopsy results. "-": without AD pathology; "+": with AD pathology; "?": the autopsy data are missing

| Case | Pathology | Observed Path | Disease Type |
|---|---|---|---|
| (1) | - | Disease-free → Dementia→ Death | Type 1 |
| (2) | - | Disease-free → Death | Unknown |
| (3) | - | Dementia→ Death | Type 1 |
| (4) | + | Disease-free → Dementia→ Death | Type 2 |
| (5) | + | Disease-free → Death | Type 2 |
| (6) | + | Dementia→ Death | Type 2 |
| (7) | ? | Disease-free → Dementia→ Death | Unknown |
| (8) | ? | Disease-free → Death | Unknown |
| (9) | ? | Dementia→ Death | Unknown |
| (10) | ? | Disease-free → Dementia | Unknown |
| (11) | ? | Disease-free | Unknown |
| (12) | ? | Dementia | Unknown |

Table A.3: Expressions of the likelihood function by observed transition paths and the autopsy results

| Case | Likelihood contribution for type 1, $\mathcal{L}^{(1)}$ | Likelihood contribution for type 2, $\mathcal{L}^{(2)}$ |
|---|---|---|
| (1) | $\psi_1\phi_1^{(1)}(a_1)p_{11}^{(1)}(a_1,a_2)p_{12}^{(1)}(a_2,a_3)p_{22}^{(1)}(a_3,a_5)\lambda_{23}^{(1)}$ | $0$ |
| (2) | $\psi_1\phi_1^{(1)}(a_1)\left[p_{11}^{(1)}(a_1,a_5)\lambda_{13}^{(1)}+p_{11}^{(1)}(a_1,a_2)p_{12}^{(1)}(a_2,a_5)\lambda_{23}^{(1)}\right]$ | $\psi_2\phi_1^{(2)}(a_1)p_{11}^{(2)}(a_1,a_5)\lambda_{14}^{(2)}$ |
| (3) | $\psi_1\phi_2^{(1)}(a_3)p_{22}^{(1)}(a_3,a_5)\lambda_{23}^{(1)}$ | $0$ |
| (4) | $0$ | $\psi_2\phi_1^{(2)}(a_1)\left[p_{12}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}\right.$ $\left.+p_{11}^{(2)}(a_1,a_2)p_{13}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}\right]+$ $\psi_2\phi_2^{(2)}(a_1)p_{22}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}$ |
| (5) | $0$ | $\psi_2\phi_1^{(2)}(a_1)\left[p_{12}^{(2)}(a_1,a_5)\lambda_{25}^{(2)}+p_{12}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_5)\lambda_{35}^{(2)}\right.$ $\left.+p_{11}^{(2)}(a_1,a_2)p_{13}^{(2)}(a_2,a_5)\lambda_{35}^{(2)}\right]+$ $\psi_2\phi_2^{(2)}(a_1)\left[p_{22}^{(2)}(a_1,a_5)\lambda_{25}^{(2)}+p_{22}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_5)\lambda_{35}^{(2)}\right]$ |
| (6) | $0$ | $\psi_2\phi_3^{(2)}(a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}$ |
| (7) | $\psi_1\phi_1^{(1)}(a_1)p_{11}^{(1)}(a_1,a_2)p_{12}^{(1)}(a_2,a_3)p_{22}^{(1)}(a_3,a_5)\lambda_{23}^{(1)}$ | $\psi_2\phi_1^{(2)}(a_1)\left[p_{12}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}\right.$ $\left.+p_{11}^{(2)}(a_1,a_2)p_{13}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}\right]+$ $\psi_2\phi_2^{(2)}(a_1)p_{22}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}$ |
| (8) | $\psi_1\phi_1^{(1)}(a_1)p_{11}^{(1)}(a_1,a_5)\lambda_{13}^{(1)}+p_{11}^{(1)}(a_1,a_2)p_{12}^{(1)}(a_2,a_5)\lambda_{23}^{(1)}$ | $\psi_2\phi_1^{(2)}(a_1)\left[p_{11}^{(2)}(a_1,a_5)\lambda_{14}^{(2)}+p_{11}^{(2)}(a_1,a_2)p_{13}^{(2)}(a_2,a_5)\lambda_{35}^{(2)}\right.$ $\left.+p_{12}^{(2)}(a_1,a_5)\lambda_{25}^{(2)}+p_{12}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_5)\lambda_{35}^{(2)}\right]+$ $\psi_2\phi_2^{(2)}(a_1)\left[p_{22}^{(2)}(a_1,a_5)\lambda_{25}^{(2)}+p_{22}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_5)\lambda_{35}^{(2)}\right]$ |
| (9) | $\psi_1\phi_2^{(1)}(a_3)p_{22}^{(1)}(a_3,a_5)\lambda_{23}^{(1)}$ | $\psi_2\phi_3^{(2)}(a_3)p_{33}^{(2)}(a_3,a_5)\lambda_{35}^{(2)}$ |
| (10) | $\psi_1\phi_1^{(1)}(a_1)p_{11}^{(1)}(a_1,a_2)p_{12}^{(1)}(a_2,a_3)p_{22}^{(1)}(a_3,a_4)$ | $\psi_2\phi_1^{(2)}(a_1)\left[p_{12}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_4)\right.$ $\left.+p_{11}^{(2)}(a_1,a_2)p_{13}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_4)\right]+$ $\psi_2\phi_2^{(2)}(a_1)p_{22}^{(2)}(a_1,a_2)p_{23}^{(2)}(a_2,a_3)p_{33}^{(2)}(a_3,a_4)$ |
| (11) | $\psi_1\phi_1^{(1)}(a_1)p_{11}^{(1)}(a_1,a_2)$ | $\psi_2\phi_1^{(2)}(a_1)\left[p_{11}^{(2)}(a_1,a_2)+p_{12}^{(2)}(a_1,a_2)\right]+\psi_2\phi_2^{(2)}(a_1)p_{22}^{(2)}(a_1,a_2)$ |
| (12) | $\psi_1\phi_2^{(1)}(a_3)p_{22}^{(1)}(a_3,a_4)$ | $\psi_2\phi_3^{(2)}(a_3)p_{33}^{(2)}(a_3,a_4)$ |

### A.3.1 Closed Forms for Disease Type One

The parameters used in the following formulas, $\lambda_{ij}$, $h_j$ and $\pi_j$, are the corresponding parameters for disease type one, i.e. $\lambda_{ij}^{(1)}$, $h_j^{(1)}$ and $\pi_j^{(1)}$, for $i, j \in \mathcal{J}^{(1)}$. For disease type one, $\mathcal{L}_7^{(1)} = \mathcal{L}_1^{(1)}$, $\mathcal{L}_8^{(1)} = \mathcal{L}_2^{(1)}$,$\mathcal{L}_9^{(1)} = \mathcal{L}_3^{(1)}$ and $\mathcal{L}_4^{(1)} = \mathcal{L}_5^{(1)} = \mathcal{L}_6^{(1)} = 0$.

$$
\begin{aligned}
\mathcal{L}_1^{(1)} &= \begin{cases} \psi_1\phi_1(a_0)\frac{-\lambda_{12}\lambda_{23}}{h_1-\lambda_{23}}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{23}a_{35}} - e^{-h_1a_{12}-\lambda_{23}(a_{23}+a_{35})}\right], & \text{if } h_1 \neq \lambda_{23} \\ \psi_1\phi_1(a_0)\lambda_{12}\lambda_{23}a_{23}e^{-\lambda_{23}(a_{12}+a_{23}+a_{35})}, & \text{if } h_1 = \lambda_{23} \end{cases} \\
\mathcal{L}_2^{(1)} &= \begin{cases} \psi_1\phi_1(a_0)\left\{\lambda_{13}e^{-h_1a_{15}} - \frac{\lambda_{12}\lambda_{23}}{h_1-\lambda_{23}}\left[e^{-h_1(a_{12}+a_{25})} - e^{-h_1a_{12}-\lambda_{23}a_{25}}\right]\right\}, & \text{if } h_1 \neq \lambda_{23} \\ \psi_1\phi_1(a_0)\left\{\lambda_{13}e^{-h_1a_{15}} + \lambda_{12}\lambda_{23}a_{25}e^{-\lambda_{23}(a_{12}+a_{25})}\right\}, & \text{if } h_1 = \lambda_{23} \end{cases} \\
\mathcal{L}_3^{(1)} &= \psi_1\phi_2(a_0)\lambda_{23}e^{-\lambda_{23}a_{35}} \\
\mathcal{L}_{10}^{(1)} &= \begin{cases} \psi_1\phi_1(a_0)\frac{-\lambda_{12}}{h_1-\lambda_{23}}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{23}a_{34}} - e^{-h_1a_{12}-\lambda_{23}(a_{23}+a_{34})}\right], & \text{if } h_1 \neq \lambda_{23} \\ \psi_1\phi_1(a_0)\lambda_{12}a_{23}e^{-\lambda_{23}(a_{12}+a_{23}+a_{34})}, & \text{if } h_1 = \lambda_{23} \end{cases} \\
\mathcal{L}_{11}^{(1)} &= \psi_1\phi_1(a_0)e^{-h_1a_{12}} \\
\mathcal{L}_{12}^{(1)} &= \psi_1\phi_2(a_0)e^{-\lambda_{23}a_{34}}
\end{aligned}
$$

### A.3.2 Closed Forms for Disease Type Two

The parameters used in the following formulas, $\lambda_{ij}$, $h_j$ and $\pi_j$, are the corresponding parameters for disease type two, i.e. $\lambda_{ij}^{(1)}$, $h_j^{(2)}$ and $\pi_j^{(2)}$, for $i, j \in \mathcal{J}^{(2)}$. For all of the cases in disease type two, $\mathcal{L}_7^{(2)} = \mathcal{L}_4^{(2)}, \mathcal{L}_8^{(2)} = \mathcal{L}_2^{(2)} + \mathcal{L}_5^{(2)}, \mathcal{L}_9^{(2)} = \mathcal{L}_6^{(2)}$ and $\mathcal{L}_1^{(2)} = \mathcal{L}_3^{(2)} = 0$.

If $h_1 \neq h_2 \neq \lambda_{35}$,

$$
\begin{aligned}
\mathcal{L}_2^{(2)} &= \psi_2\phi_1(a_1)\lambda_{14}e^{-h_1 a_{15}} \\
\mathcal{L}_4^{(2)} &= \psi_2\phi_1(a_1)\left\{\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-h_2)(h_1-\lambda_{35})}e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}} + \frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-\lambda_{35})(h_2-\lambda_{35})}e^{-h_1 a_{12}-\lambda_{35}(a_{23}+a_{35})}\right. \\
&\quad \left. -\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-h_2)(h_2-\lambda_{35})}\left[e^{-h_1 a_{12}-\lambda_{35}(a_{23}+a_{35})} + e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_2 a_{12}-\lambda_{35}(a_{23}+a_{35})}\right]\right\} \\
&\quad -\psi_2\phi_2(a_1)\frac{\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}\left[e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_2 a_{12}-\lambda_{35}(a_{23}+a_{35})}\right] \\
\mathcal{L}_5^{(2)} &= \psi_2\phi_1(a_1)\left\{-\frac{\lambda_{12}\lambda_{25}}{h_1-h_2}\left[e^{-h_1 a_{15}} - e^{-h_2 a_{15}}\right] + \frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-h_2)(h_1-\lambda_{35})}e^{-h_1(a_{12}+a_{25})}\right. \\
&\quad \left. -\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-h_2)(h_2-\lambda_{35})}\left[e^{-h_2(a_{12}+a_{25})} + e^{-h_1 a_{12}-\lambda_{35}a_{25}} - e^{-h_2 a_{12}-\lambda_{35}a_{25}}\right] + \frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-\lambda_{35})(h_2-\lambda_{35})}e^{-h_1 a_{12}-\lambda_{35}a_{25}}\right\} \\
&\quad +\psi_2\phi_2(a_1)\left\{\lambda_{25}e^{-h_2 a_{15}} - \frac{\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}\left[e^{-h_2(a_{12}+a_{25})} - e^{-h_2 a_{12}-\lambda_{35}a_{25}}\right]\right\} \\
\mathcal{L}_6^{(2)} &= \psi_2\phi_3(a_3)\lambda_{35}e^{-\lambda_{35}a_{35}} \\
\mathcal{L}_{10}^{(2)} &= \psi_2\phi_1(a_1)\left\{\frac{\lambda_{12}\lambda_{23}}{(h_1-h_2)(h_1-\lambda_{35})}e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{34}} + \frac{\lambda_{12}\lambda_{23}}{(h_1-\lambda_{35})(h_2-\lambda_{35})}e^{-h_1 a_{12}-\lambda_{35}(a_{23}+a_{34})}\right. \\
&\quad \left. -\frac{\lambda_{12}\lambda_{23}}{(h_1-h_2)(h_2-\lambda_{35})}\left[e^{-h_1 a_{12}-\lambda_{35}(a_{23}+a_{34})} + e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{34}} - e^{-h_2 a_{12}-\lambda_{35}(a_{23}+a_{34})}\right]\right\} \\
&\quad -\psi_2\phi_2(a_1)\frac{\lambda_{23}}{h_2-\lambda_{35}}\left[e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{34}} - e^{-h_2 a_{12}-\lambda_{35}(a_{23}+a_{34})}\right] \\
\mathcal{L}_{11}^{(2)} &= \psi_2\phi_1(a_1)\left\{e^{-h_1 a_{12}} - \frac{\lambda_{12}}{h_1-h_2}\left[e^{-h_1 a_{12}} - e^{-h_2 a_{12}}\right]\right\} + \psi_2\phi_2(a_1)e^{-h_2 a_{12}} \\
\mathcal{L}_{12}^{(2)} &= \psi_2\phi_3(a_3)e^{-\lambda_{35}a_{34}}
\end{aligned}
$$

If $h_1 = h_2 \neq \lambda_{35}$,

$$
\begin{aligned}
\mathcal{L}_2^{(2)} &= \psi_2\phi_1(a_1)\lambda_{14}e^{-h_1a_{15}} \\
\mathcal{L}_4^{(2)} &= \psi_2\phi_1(a_1)\left\{-\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_1-\lambda_{35}}\left(\left(a_{12}-a_{23}\right)e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}} - a_{12}e^{-h_1a_{12}-\lambda_{35}(a_{23}+a_{35})}\right)\right. \\
&\quad \left.-\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{\left(h_1-\lambda_{35}\right)^2}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_1a_{12}-\lambda_{35}(a_{23}+a_{35})}\right]\right\} \\
&\quad -\psi_2\phi_2(a_1)\frac{\lambda_{23}\lambda_{35}}{h_1-\lambda_{35}}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_1a_{12}-\lambda_{35}(a_{23}+a_{35})}\right] \\
\mathcal{L}_5^{(2)} &= \psi_2\phi_1(a_1)\left\{\lambda_{12}\lambda_{25}a_{15}e^{-h_1a_{15}} - \frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_1-\lambda_{35}}\left(\left(a_{12}-a_{25}\right)e^{-h_1(a_{12}+a_{25})} - a_{12}e^{-h_1a_{12}-\lambda_{35}a_{25}}\right)\right. \\
&\quad \left.-\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{\left(h_1-\lambda_{35}\right)^2}\left[e^{-h_1(a_{12}+a_{25})} - e^{-h_1a_{12}-\lambda_{35}a_{25}}\right]\right\} \\
&\quad +\psi_2\phi_2(a_1)\left\{\lambda_{25}e^{-h_1a_{15}} - \frac{\lambda_{23}\lambda_{35}}{h_1-\lambda_{35}}\left[e^{-h_1(a_{12}+a_{25})} - e^{-h_1a_{12}-\lambda_{35}a_{25}}\right]\right\} \\
\mathcal{L}_6^{(2)} &= \psi_2\phi_3(a_1)\lambda_{35}e^{-\lambda_{35}a_{35}} \\
\mathcal{L}_{10}^{(2)} &= \psi_2\phi_1(a_1)\left\{\frac{\lambda_{12}\lambda_{23}}{h_1-\lambda_{35}}\left(-\left(a_{12}-a_{23}\right)e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{34}} + a_{12}e^{-h_1a_{12}-\lambda_{35}(a_{23}+a_{34})}\right)\right. \\
&\quad \left.-\frac{\lambda_{12}\lambda_{23}}{\left(h_1-\lambda_{35}\right)^2}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{34}} - e^{-h_1a_{12}-\lambda_{35}(a_{23}+a_{34})}\right]\right\} \\
&\quad -\psi_2\phi_2(a_1)\frac{\lambda_{23}}{h_1-\lambda_{35}}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_1a_{12}-\lambda_{35}(a_{23}+a_{34})}\right] \\
\mathcal{L}_{11}^{(2)} &= \psi_2\phi_1(a_1)\left\{e^{-h_1a_{12}} + \lambda_{12}a_{12}e^{-h_1a_{12}}\right\} + \psi_2\phi_2(a_1)e^{-h_1a_{12}} \\
\mathcal{L}_{12}^{(2)} &= \psi_2\phi_3(a_1)e^{-\lambda_{35}a_{34}}
\end{aligned}
$$

If $h_2 \neq h_1 = \lambda_{35}$,

$$
\begin{aligned}
\mathcal{L}_2^{(2)} &= \psi_2\phi_1(a_1)\lambda_{14}e^{-\lambda_{35}a_{15}} \\
\mathcal{L}_4^{(2)} &= \psi_2\phi_1(a_1)\left\{\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_2-\lambda_{35})^2}\left[e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_2a_{12}-\lambda_{35}(a_{23}+a_{35})}\right] + \frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}a_{23}e^{-\lambda_{35}(a_{12}+a_{23}+a_{35})}\right\} \\
&\quad -\psi_2\phi_2(a_1)\frac{\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}\left[e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{35}} - e^{-h_2a_{12}-\lambda_{35}(a_{23}+a_{35})}\right] \\
\mathcal{L}_5^{(2)} &= \psi_2\phi_1(a_1)\left\{\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_2-\lambda_{35})^2}\left[e^{-\lambda_{35}a_{12}-h_2a_{25}} - 2e^{-\lambda_{35}(a_{12}+a_{25})} - e^{-h_2(a_{12}+a_{25})} + 2e^{-h_2a_{25}-\lambda_{35}a_{12}}\right]\right. \\
&\quad \left. -\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}e^{-\lambda_{35}(a_{12}+a_{25})} - \frac{\lambda_{12}\lambda_{25}}{h_2-\lambda_{35}}\left[e^{-h_2a_{15}} - e^{-\lambda_{35}a_{15}}\right]\right\} \\
&\quad +\psi_2\phi_2(a_1)\left\{\lambda_{25}e^{-h_2a_{15}} - \frac{\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}\left[e^{-h_2(a_{12}+a_{25})} - e^{-h_2a_{12}-\lambda_{35}a_{25}}\right]\right\} \\
\mathcal{L}_6^{(2)} &= \psi_2\phi_3(a_1)\lambda_{35}e^{-\lambda_{35}a_{35}} \\
\mathcal{L}_{10}^{(2)} &= \psi_2\phi_1(a_1)\left\{\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_2-\lambda_{35})^2}\left[e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{34}} - e^{-h_2a_{12}-\lambda_{35}(a_{23}+a_{34})}\right] + \frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}a_{23}e^{-\lambda_{35}(a_{12}+a_{23}+a_{34})}\right\} \\
&\quad -\psi_2\phi_2(a_1)\frac{\lambda_{23}\lambda_{35}}{h_2-\lambda_{35}}\left[e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{34}} - e^{-h_2a_{12}-\lambda_{35}(a_{23}+a_{34})}\right] \\
\mathcal{L}_{11}^{(2)} &= \psi_2\phi_1(a_1)\left\{e^{-\lambda_{35}a_{12}} - \frac{\lambda_{12}}{h_2-\lambda_{35}}\left[e^{-h_2a_{12}} - e^{-\lambda_{35}a_{12}}\right]\right\} + \psi_2\phi_2(a_1)e^{-h_2a_{12}} \\
\mathcal{L}_{12}^{(2)} &= \psi_2\phi_3(a_1)e^{-\lambda_{35}a_{34}}
\end{aligned}
$$

If $h_1 \neq h_2 = \lambda_{35}$,

$$
\begin{aligned}
\mathcal{L}_2^{(2)} &= \psi_2\phi_1(a_1)\lambda_{14}e^{-h_1 a_{15}} \\
\mathcal{L}_4^{(2)} &= \psi_2\phi_1(a_1)\left\{-\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_1-\lambda_{35}}a_{23}e^{-\lambda_{35}(a_{12}+a_{23}+a_{35})}\right. \\
&\quad \left.-\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-\lambda_{35})^2}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}}-e^{-h_1 a_{12}-\lambda_{35}(a_{23}+a_{35})}\right]\right\} \\
&\quad +\psi_2\phi_2(a_1)\lambda_{23}\lambda_{35}a_{23}e^{-\lambda_{35}(a_{12}+a_{23}+a_{35})} \\
\mathcal{L}_5^{(2)} &= \psi_2\phi_1(a_1)\left\{-\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{h_1-\lambda_{35}}\left((a_{23}-a_{25})\,e^{-h_1 a_{12}-\lambda_{35}a_{25}}-a_{23}e^{-\lambda_{35}(a_{12}+a_{25})}\right)\right. \\
&\quad \left.+\frac{\lambda_{12}\lambda_{23}\lambda_{35}}{(h_1-\lambda_{35})^2}\left[e^{-h_1(a_{12}+a_{25})}-e^{-h_1 a_{12}-\lambda_{35}a_{25}}\right]-\frac{\lambda_{12}\lambda_{25}}{h_1-\lambda_{35}}\left[e^{-h_1 a_{15}}-e^{-\lambda_{35}a_{15}}\right]\right\} \\
&\quad +\psi_2\phi_2(a_1)\left\{\lambda_{25}e^{-\lambda_{35}a_{15}}+\lambda_{23}\lambda_{35}a_{25}e^{-\lambda_{35}(a_{12}+a_{25})}\right\} \\
\mathcal{L}_6^{(2)} &= \psi_2\phi_3(a_1)\lambda_{35}e^{-\lambda_{35}a_{35}} \\
\mathcal{L}_{10}^{(2)} &= \psi_2\phi_1(a_1)\left\{-\frac{\lambda_{12}\lambda_{23}}{h_1-\lambda_{35}}a_{23}e^{-\lambda_{35}(a_{12}+a_{23}+a_{34})}\right. \\
&\quad \left.-\frac{\lambda_{12}\lambda_{23}}{(h_1-\lambda_{35})^2}\left[e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{34}}-e^{-h_1 a_{12}-\lambda_{35}(a_{23}+a_{34})}\right]\right\} \\
&\quad +\psi_2\phi_2(a_1)\lambda_{23}\lambda_{35}a_{23}e^{-\lambda_{35}(a_{12}+a_{23}+a_{34})} \\
\mathcal{L}_{11}^{(2)} &= \psi_2\phi_1(a_1)\left\{e^{-h_1 a_{12}}-\frac{\lambda_{12}}{h_1-\lambda_{35}}\left[e^{-h_1 a_{12}}-e^{-\lambda_{35}a_{12}}\right]\right\}+\psi_2\phi_2(a_1)e^{-\lambda_{35}a_{12}} \\
\mathcal{L}_{12}^{(2)} &= \psi_2\phi_3(a_1)e^{-\lambda_{35}a_{34}}
\end{aligned}
$$

If $h_2 = h_1 = \lambda_{35}$,

$$
\begin{aligned}
\mathcal{L}_2^{(2)} &= \psi_2\phi_1(a_1)\lambda_{14}e^{-h_1 a_{15}} \\
\mathcal{L}_4^{(2)} &= \psi_2\phi_1(a_1)\lambda_{12}\lambda_{23}\lambda_{35}\left(a_{12}a_{23}e^{-h_1 a_{12}-h_2 a_{23}-\lambda_{35}a_{35}} + \frac{a_{23}^2}{2}e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{35}}\right) \\
&\quad +\psi_2\phi_2(a_1)\lambda_{23}\lambda_{35}a_{23}e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{35}} \\
\mathcal{L}_5^{(2)} &= \psi_2\phi_1(a_1)\left\{\lambda_{12}\lambda_{25}a_{15}e^{-h_1 a_{15}} + \lambda_{12}\lambda_{23}\lambda_{35}\left(a_{12}a_{25}e^{-h_1 a_{12}-h_2 a_{25}} + \frac{a_{25}^2}{2}e^{-h_1(a_{12}+a_{25})}\right)\right\} \\
&\quad +\psi_2\phi_2(a_1)\left\{\lambda_{25}e^{-h_2 a_{15}} + \lambda_{23}\lambda_{35}a_{25}e^{-h_2(a_{12}+a_{25})}\right\} \\
\mathcal{L}_6^{(2)} &= \psi_2\phi_3(a_1)\lambda_{35}e^{-\lambda_{35}a_{35}} \\
\mathcal{L}_{10}^{(2)} &= \psi_2\phi_1(a_1)\lambda_{12}\lambda_{23}\left(a_{12}a_{23}e^{-h_1 a_{12}-h_2 a_{23}-\lambda_{35}a_{34}} + \frac{a_{23}^2}{2}e^{-h_1(a_{12}+a_{23})-\lambda_{35}a_{34}}\right) \\
&\quad +\psi_2\phi_2(a_1)\lambda_{23}a_{23}e^{-h_2(a_{12}+a_{23})-\lambda_{35}a_{34}} \\
\mathcal{L}_{11}^{(2)} &= \psi_2\phi_1(a_1)\left\{e^{-h_1 a_{12}} + \lambda_{12}a_{12}e^{-h_1 a_{12}}\right\} + \psi_2\phi_2(a_1)e^{-h_2 a_{12}} \\
\mathcal{L}_{12}^{(2)} &= \psi_2\phi_3(a_1)e^{-\lambda_{35}a_{34}}
\end{aligned}
$$

# Appendix B

# Appendix for Chapter 4

## B.1 Complementary Analysis Results

Table B.1 gives the analyses results from the Nun Study where the effects of $APOE-\epsilon 4$ on the mortality transitions are not fixed at 0. The results are based on a piecewise constant intensity model with a breakpoint at age 90. According to the estimates and standard errors for $\beta_{ij}$'s on the transitions into the death state from NC, MCI and dementia, the p-values for these parameters can be calculated as 0.268, 0.722 and 0.898 respectively. Therefore, we ignored the effects on the mortality transitions for the analyses reported in Table 4.2.

Table B.1: Results from the Nun Study using a piecewise constant intensity model, with a breakpoint at age 90, considering effects of $APOE - \epsilon 4$ on mortality. RV: relative reduction in variance of the MLE from $\mathcal{L}$ versus $\mathcal{L}_0$. $\alpha_{ij}$: the log ratio of rates for age 90+ for transition from state $i$ to state $j$; $\beta_{ij}$: the log ratio of rates for having $APOE - \epsilon 4$ for transition from state $i$ to state $j$, where $\beta_{i4}, i \in \mathcal{J}$ were fixed at 0.

| | | Conditional, $\mathcal{L}_0$ | | Weighted, $\mathcal{L}$ | | |
|---|---|---|---|---|---|---|
| | Transitions | EST | SE | EST | SE | $RV$ % |
| Log Baseline | NC to MCI | -1.300 | 0.079 | -1.141 | 0.064 | 34.637 |
| 75-90 | NC to Death | -3.936 | 0.328 | -3.947 | 0.348 | -12.634 |
| $\log \lambda_{ij0}$ | MCI to NC | -1.707 | 0.088 | -1.876 | 0.090 | -5.350 |
| | MCI to Dementia | -2.781 | 0.123 | -2.937 | 0.109 | 21.341 |
| | MCI to Death | -2.734 | 0.131 | -2.781 | 0.127 | 5.227 |
| | Dementia to Death | -1.742 | 0.115 | -1.696 | 0.104 | 18.041 |
| Log RR, $X = 1$ | NC to MCI | 0.186 | 0.198 | 0.117 | 0.151 | 41.464 |
| $\beta_{ij}$ | NC to Death | 0.666 | 0.600 | 0.675 | 0.609 | -2.963 |
| | MCI to NC | -0.937 | 0.260 | -0.852 | 0.261 | -0.669 |
| | MCI to Dementia | 0.524 | 0.185 | 0.592 | 0.155 | 29.962 |
| | MCI to Death | 0.040 | 0.241 | 0.085 | 0.238 | 3.056 |
| | Dementia to Death | 0.038 | 0.131 | 0.016 | 0.124 | 9.625 |
| Log RR, $\geq$90 | NC to MCI | 0.511 | 0.166 | 0.494 | 0.147 | 21.655 |
| $\alpha_{ij}$ | NC to Death | 1.539 | 0.456 | 1.654 | 0.470 | -5.971 |
| | MCI to NC | -0.881 | 0.232 | -0.811 | 0.235 | -2.437 |
| | MCI to Dementia | 0.951 | 0.169 | 1.131 | 0.153 | 17.749 |
| | MCI to Death | 0.811 | 0.183 | 0.838 | 0.181 | 2.111 |
| | Dementia to Death | 0.831 | 0.128 | 0.771 | 0.119 | 13.092 |

## B.2 Expressions of the Transition Probabilities

Consider a time-homogeneous Markov model as shown in Figure 4.1. The transition intensity matrix is in the form of

$$\Lambda = \begin{bmatrix} -h_1 & \lambda_{12} & 0 & \lambda_{14} \\ \lambda_{21} & -h_2 & \lambda_{23} & \lambda_{24} \\ 0 & 0 & -h_3 & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $h_i = \sum_{j=1}^{4} \lambda_{ij}$. The transition probability matrix is $\mathrm{P}(s,t) = e^{(t-s)\Lambda}$. Consider the eigendecomposition of the intensity matrix

$$\Lambda = \mathrm{VDV}^{-1},$$

where D is a diagonal matrix with elements being the eigenvalues of $\Lambda$, and the columns of V are the eigenvectors of $\Lambda$ that correspond to elements of D. Then the probability matrix is

$$\mathrm{P}(s,t) = \mathrm{V}e^{(t-s)\mathrm{D}}\mathrm{V}^{-1}. \tag{B.1}$$

Denote

$$\begin{aligned} a_1 &= -\frac{1}{2}\left(h_1 + h_2 + \sqrt{(h_1-h_2)^2 + 4\lambda_{12}\lambda_{21}}\right) \\ a_2 &= -\frac{1}{2}\left(h_1 + h_2 - \sqrt{(h_1-h_2)^2 + 4\lambda_{12}\lambda_{21}}\right) \\ b &= h_1h_2 - h_1h_3 - h_2h_3 + h_3^2 - \lambda_{12}\lambda_{21}. \end{aligned}$$

The closed forms for the elements of the transition probability matrix are the following.

$$
\begin{aligned}
P_{11}(s,t) &= \frac{(a_1+h_2)e^{a_1(t-s)}-(a_2+h_2)e^{a_2(t-s)}}{a_1-a_2} \\
P_{12}(s,t) &= \frac{\lambda_{21}\left[e^{a_2(t-s)}-e^{a_1(t-s)}\right]}{a_1-a_2} \\
P_{13}(s,t) &= \frac{\lambda_{12}\lambda_{23}}{b}e^{-\lambda_{34}(t-s)} - \frac{\lambda_{12}\lambda_{23}}{(a_1-a_2)b}\left[(a_1+h_1+h_2-\lambda_{34})\,e^{a_1(t-s)}-(a_2+h_1+h_2-\lambda_{34})\,e^{a_2(t-s)}\right] \\
P_{14}(s,t) &= 1-\sum_{k=1}^{3}P_{1k}(s,t) \\
P_{21}(s,t) &= \frac{\lambda_{12}\left[e^{a_2(t-s)}-e^{a_1(t-s)}\right]}{a_1-a_2} \\
P_{22}(s,t) &= \frac{(a_2+h_2)e^{a_1(t-s)}-(a_1+h_2)e^{a_2(t-s)}}{a_1-a_2} \\
P_{23}(s,t) &= -\frac{\lambda_{23}(h_1-\lambda_{34})}{b}e^{-\lambda_{34}(t-s)} \\
&\quad +\frac{\lambda_{23}}{(a_1-a_2)b}\left\{\left[\lambda_{12}^2-(h_1-\lambda_{34})(a_2+h_2)\right]e^{a_1(t-s)}-\left[\lambda_{12}^2-(h_1-\lambda_{34})(a_1+h_2)\right]e^{a_2(t-s)}\right\} \\
P_{24}(s,t) &= 1-\sum_{k=1}^{3}P_{2k}(s,t) \\
P_{33}(s,t) &= e^{-\lambda_{34}(t-s)} \\
P_{34}(s,t) &= 1-e^{-\lambda_{34}(t-s)} \\
P_{44}(s,t) &= 1 \\
0 &= P_{31}(s,t)=P_{32}(s,t)=P_{41}(s,t)=P_{42}(s,t)=P_{43}(s,t)
\end{aligned}
$$

## B.3 Reborn Equilibrium Probabilities for Multistate Models with Absorbing States

Suppose $Z(t)$ is a Markov model with state space $\mathcal{J} = \{1, 2, \ldots, J\}$ and transition intensity matrix $\Lambda$, where $\mathcal{J}$ contains at least one absorbing state. Let $\mathcal{J}_A$ and $\mathcal{J}_B$ be the sets of absorbing states and transient states respectively, $J_A$ and $J_B$ the numbers of absorbing and transient states, so that $J = J_A + J_B$. Such a model does not have an equilibrium or a stationary distribution because all of the subjects will end up in one of the absorbing states given enough time.

However, the equilibrium distribution can be obtained for a reborn process where individuals are assumed to go back to a transient state immediately after they enter into any absorbing state. Without loss of generality, let $\mathcal{J}_{\mathcal{B}} = \{1, 2, \ldots, J_B\}$, and $\mathcal{J}_{\mathcal{A}} = \{J_B + 1, J_B + 2, \ldots, J\}$. Let $\omega_{ab}$ be the probability of going from an absorbing state $a$ into a transient state $b$, so that $\sum_{b \in \mathcal{J}_B} \omega_{ab} = 1, \forall a \in \mathcal{J}_A$. Such a process is equivalent to a Markov model $Z_B(t)$ with state space $\mathcal{J}_B$ and transition intensity matrix

$$\Lambda_B \;=\; \mathrm{L}\Lambda\mathrm{R}, \tag{B.2}$$

where L is a $J_B \times J$ matrix constructed by the first $J_B$ rows of an $J-$dimensional identity matrix, and

$$\mathrm{R} \;=\; \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \omega_{J_B+1,1} & \cdots & \cdots & \omega_{J_B+1,J_B} \\ \vdots & & & \vdots \\ \omega_{J,1} & \cdots & \cdots & \omega_{J,J_B} \end{bmatrix}_{J \times J_B.} \tag{B.3}$$

Let a row vector $\phi = (\phi_1, \phi_2, ..., \phi_{J_B})$ be equilibrium probability of the reborn process

$Z_B(t)$. The values of $\phi$ can be obtained by solving

$$\phi\Lambda_B \;=\; \mathbf{0}, \tag{B.4}$$

subject to $\sum_{j=1}^{J_B}\phi_j = 1$ and $0 \le \phi_j \le 1, j \in \mathcal{J}_B$. If we want to use the equilibrium probability as the auxiliary prevalence data for a piecewise constant model, we need to identify that which time interval does the recruitment time fall in, and the equilibrium probability for the reborn process can be calculated using the transition intensities for this interval.

### B.3.1 A Reborn Process for the Nun Study

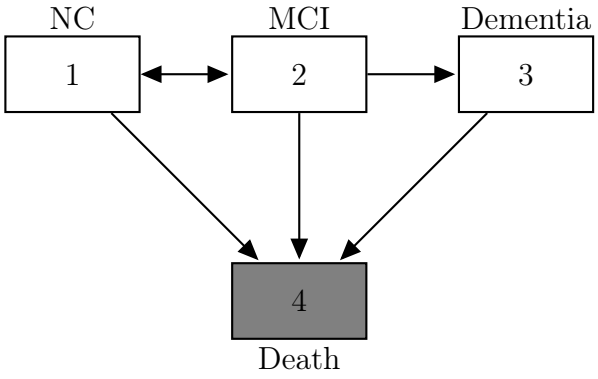


Figure B.1: A four-state stochastic process with reversible transition from mild cognitive impairment (MCI) to normal cognition (NC). The shaded block corresponds to the absorbing state.

Consider the four-state model as shown in Figure B.1, with the transition intensity matrix being

$$\Lambda \;=\; \begin{bmatrix} -(\lambda_{12}+\lambda_{14}) & \lambda_{12} & 0 & \lambda_{14} \\ \lambda_{21} & -(\lambda_{21}+\lambda_{23}+\lambda_{24}) & \lambda_{23} & \lambda_{24} \\ 0 & 0 & \lambda_{34} & -\lambda_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Assume that if a subject enters the death state, it will be reborn immediately to state $j$ with probability $\omega_j, j = 1, 2, 3$. The transition intensity matrix for the reborn process is:

$$\begin{aligned}
\Lambda_B &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \Lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \omega_1 & \omega_2 & \omega_3 \end{bmatrix} \\
&= \begin{bmatrix} -\left(\lambda_{12} + \left(\omega_2 + \omega_3\right)\lambda_{14}\right) & \lambda_{12} + \omega_2\lambda_{14} & \omega_3\lambda_{14} \\ \lambda_{21} + \omega_1\lambda_{24} & -\left(\lambda_{21} + \lambda_{23} + \left(\omega_1 + \omega_3\right)\lambda_{24}\right) & \lambda_{23} + \omega_3\lambda_{24} \\ \omega_1\lambda_{34} & \omega_2\lambda_{34} & -\left(\omega_1 + \omega_2\right)\lambda_{34} \end{bmatrix}
\end{aligned}$$

Consider a simple case where every individual is reborn in state 1; the equilibrium probabilities are:

$$\begin{aligned}
\phi_1 &= \frac{\lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}} \\
\phi_2 &= \frac{\lambda_{12}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}} \\
\phi_3 &= \frac{\lambda_{12}\lambda_{23}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{21}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}}.
\end{aligned}$$

### B.3.2 Some Examples of Reborn Processes for Alternate Models for the Nun Study

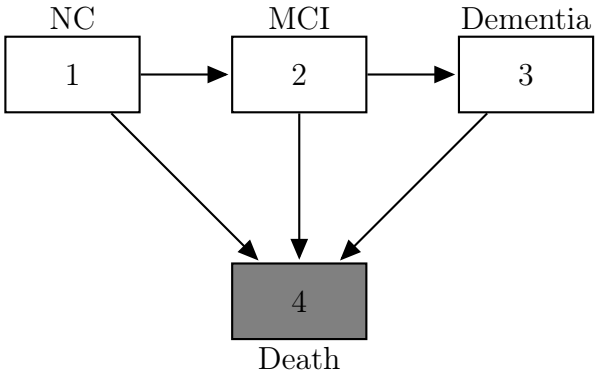


Figure B.2: A four-state stochastic process without reversible transition. The shaded block corresponds to the absorbing state.

In this section, we give two other examples of reborn processes for four-state Markov models which have an absorbing state. Equilibrium probabilities for the reborn processes will be calculated while assuming that the subjects reborn to the first state with probability 1 using a time-homogeneous Markov model.

Consider the four-state stochastic process similar to one for the Nun Study without reverse transition from MCI to NC, see Figure B.2. The transition intensity matrix for such a model is

$$\Lambda = \begin{bmatrix} -(\lambda_{12}+\lambda_{14}) & \lambda_{12} & 0 & \lambda_{14} \\ 0 & -(\lambda_{23}+\lambda_{24}) & \lambda_{23} & \lambda_{24} \\ 0 & 0 & -\lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

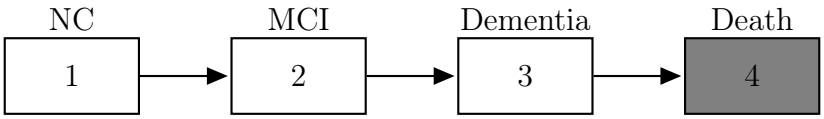


Figure B.3: A stepwise progressive four-state stochastic process. The shaded block corresponds to the absorbing state.

The equilibrium probabilities of the reborn model by solving (B.4) are

$$\begin{aligned}
\phi_1 &= \frac{\lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}} \\
\phi_2 &= \frac{\lambda_{12}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}} \\
\phi_3 &= \frac{\lambda_{12}\lambda_{23}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{23}\lambda_{34} + \lambda_{24}\lambda_{34}}.
\end{aligned}$$

Consider a stepwise progressive model, where severer stages of the disease can only be approached from the previous state, see Figure B.3. The transition intensity matrix for such a model is

$$\Lambda = \begin{bmatrix} -\lambda_{12} & \lambda_{12} & 0 & 0 \\ 0 & -\lambda_{23} & \lambda_{23} & 0 \\ 0 & 0 & -\lambda_{34} & \lambda_{34} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The equilibrium probabilities of the reborn model by solving (B.4) are

$$\begin{aligned}
\phi_1 &= \frac{\lambda_{23}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{23}\lambda_{34}} \\
\phi_2 &= \frac{\lambda_{12}\lambda_{34}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{23}\lambda_{34}} \\
\phi_3 &= \frac{\lambda_{12}\lambda_{23}}{\lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{34} + \lambda_{23}\lambda_{34}}.
\end{aligned}$$