University of Kentucky Doctoral Dissertations   Graduate School

2010

A MARKOV TRANSITION MODEL TO DEMENTIA WITH DEATH AS A COMPETING EVENT

Liou Xu
University of Kentucky, x.liou@yahoo.com

ABSTRACT OF DISSERTATION

Liou Xu

The Graduate School

University of Kentucky

2010

A MARKOV TRANSITION MODEL TO DEMENTIA
WITH DEATH AS A COMPETING EVENT

───────────────────────────────────

ABSTRACT OF DISSERTATION

───────────────────────────────────

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

By
Liou Xu

Lexington, Kentucky

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

Lexington, Kentucky

2010

ABSTRACT OF DISSERTATION

A MARKOV TRANSITION MODEL TO DEMENTIA
WITH DEATH AS A COMPETING EVENT

The research on multi-state Markov transition model is motivated by the nature of the longitudinal data from the Nun Study (Snowdon, 1997), and similar information on the BRAiNS cohort (Salazar, 2004). Our goal is to develop a flexible methodology for handling the categorical longitudinal responses and competing risks time-to-event that characterizes the features of the data for research on dementia. To do so, we treat the survival from death as a continuous variable rather than defining death as a competing absorbing state to dementia. We assume that within each subject the survival component and the Markov process are linked by a shared latent random effect, and moreover, these two pieces are conditionally independent given the random effect and their corresponding predictor variables. The problem of the dependence among observations made on the same subject (repeated measurements) is addressed by assuming a first order Markovian dependence structure.

A closed-form expression for the individual and thus overall conditional marginal likelihood function is derived, which we can evaluate numerically to produce the maximum likelihood estimates for the unknown parameters. This method can be implemented using standard statistical software such as SAS Proc Nlmixed©. We present the results of simulation studies designed to show how the model's ability to accurately estimate the parameters can be affected by the distributional form of the survival term.

Then we focus on addressing the problem by accommodating the residual life time of the subject's confounding in the nonhomogeneous chain. The convergence status of the chain is examined and the formulation of the absorption statistics is derived. We propose using the Delta method to estimate the variance terms for construction of confidence intervals. The results are illustrated with applications to the Nun Study data in details.

Liou Xu

December 8, 2010

A MARKOV TRANSITION MODEL TO DEMENTIA
WITH DEATH AS A COMPETING EVENT

By

Liou Xu

Richard J. Kryscio, Ph.D.

Director of Dissertation

Arne C. Bathke, Ph.D.

Director of Graduate Studies

December 8, 2010

Date

RULES FOR USE OF DISSERTATIONS

DISSERTATION

Liou Xu

The Graduate School

University of Kentucky

2010

A MARKOV TRANSITION MODEL TO DEMENTIA
WITH DEATH AS A COMPETING EVENT

DISSERTATION

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

By
Liou Xu

Lexington, Kentucky

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

Lexington, Kentucky

2010

ACKNOWLEDGMENTS

First, I thank my advisor Dr. Richard Kryscio for his comments, insights, guidance, and encouragement that made possible this work. Next, I am grateful with Dr. David Snowdon and Dr. Suzanne Tyas for making the Nun Study data set available. Also, I would like to thank my dissertation committee consisting of Dr. Arne Bathke, Dr. William Griffith, Dr. Cidambi Srinivasan, and Dr. Marta Mendiondo.

In addition to the people I mentioned above, I wish to thank the faculty, students, and staff from the Department of Statistics and the Sanders Brown Center on Aging at University of Kentucky for all the knowledge, help, and support I received from them. Especially, I would like to express my gratitude to those members whose classes I took: Dr. William Griffth, Dr. Arne Bathke, Dr. Cidambi Srinivasan, Dr. Mai Zhou, Dr. Kert Veile, Dr. Richard Kryscio, Dr. William Rayens, Dr. Z. Govindarajulu, and Dr. Richard Charnigo. Separately, I thank Dr. Arnold Stromberg and Dr. Constance Wood, who played signicant role in my enrollment in the University of Kentucky.

I owe my gratitude to all the people who have made this dissertation possible. It is not possible to mention all of them, and I apologize to those I have left out.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 : Introduction

1.1 Background

In longitudinal studies it is common that repeated measurements on a response, an observation on a possibly censored time-to-event, and additional covariate information are collected on each participant. In most medical research interest often focuses on modeling and interpreting the interrelationships among these variables. A familiar example is that of studies on progression to dementia, covariates including demographic information, such as age, education level, and some gene-related factor, are recorded at baseline, and the outcome variable of interest is a series of correlated categorical responses which are observed at certain time points, sometimes several years apart. Time to progression to death is also recorded for each participant, although some subjects may fail to experience the event ("dementia" or "death") by the time the study closes. The primary objectives of the study are (i) to understand within-subject patterns of transition among pre-disease states and dementia; (ii) to characterize the relationship between the risk of developing dementia across the long-term trajectory from time to death. However, addressing these objectives in practice is much more difficult depending on the nature of the data actually observed. The complications posed by the realities and the potential for biased inferences if naïve techniques are applied have led to considerable recent interest in so-called joint models, where models for the event time distribution and longitudinal data are taken to depend on some shared latent random effect. A desirable feature for joint modeling is that in the absence of the presumed association between the

longitudinal data and event times, the analysis should cover the same results as would be obtained from separate analyses for each component.

Most previous work has been based on specific applications. Hogan and Laird (1997ab) give an excellent review of models and methods for joint analysis of data of this type. A well-known application is in AIDS research in which a biomarker such as CD4 lymphocyte count is determined intermittently and its relationship with time to seroconversion or death is of interest (Pawitan and Self, 1993; Tsiatis *et al.*, 1995; Wulfsohn and Tsiatis, 1997). Follman and Wu (1995) develop a class of random effects dependent selection models in the more general setting of the shared parameter models, which can also account for missing observations. The approach uses generalized linear models

$$\zeta_Y(E[Y_i|b_i]) = X_i\beta + W_i b_i$$

$$\zeta_D(E[D_i|b_i]) = Z_i\phi_1 + \phi_2 b_i$$

where b_i represents the shared random effect for the ith subject, ζ_Y and ζ_D are monotone link functions, X_i, Z_i and W_i are fixed covariate matrices, and β, ϕ_1, and ϕ_2 are the parameters associated with X_i, Z_i and b_i, respectively. The joint distribution of Y_i and D_i

$$f(y_i, d_i) = \int_{b_i} f(y_i|b_i)f(d_i|b_i)dF(b_i)$$

is said to follow the shared random parameter model. Likelihood-based estimation procedures can accommodate right-censored values of D_i. The likelihood function for a random effects model with censored D_i is

$$L(\psi_{D|Y}, \psi_Y) = \prod_{i=1}^{N} \int f(y_i|b_i; \beta, \sigma^2) f(b_i; \Gamma) f(\tilde{d}_\iota, \tau_i|b_i; \psi_{D|Y}) db_i$$

where

$$f(\tilde{d}_\iota, \tau_i|b_i; \psi_{D|Y}) = \left[f_{D|b}(\tilde{d}_\iota|b_i; \psi_{D|Y})\right]^{\tau_i} \left[1 - F_{D|b}(\tilde{d}_\iota|b_i; \psi_Y)\right]^{1-\tau_i}.$$

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

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

The model is established on the basis of three major assumptions

(a) T and Y are conditionally independent given η

(b) X can affect T either through η or directly

(c) X only affects Y through its influence on η

To be more specific, $Y_i(t)$, the observed value of the process at time t is modeled as an independent observation from a generalized linear model (GLM) with linear predictor $\eta_i(t)$. That is

$$g[E\{Y_i(t)|\eta_i(t)\}] = \eta_i(t)$$

where $\eta_i(t)$ is generally assumed to follow a Gaussian stochastic process. And the model allows different forms of conditional hazard to be specified for $[T|\eta, X]$. An application of this model is when the auxiliary variable Y is an imperfect surrogate end point for T.

Fieuws and Verbeke (2006) propose a pairwise approach to resolve the computational complexity of high-dimensional joint random effects models. In such framework, estimates for the elements in the parameter space are obtained by maximizing each of the likelihoods of the pairwise bivariate models separately, instead of maximizing the likelihood of the joint mixed model.

Elashoff *et al*. (2007) suggest joint modeling of the repeated measurements and competing risk failure time data to allow for more than one distinct failure type in the survival endpoint. The joint model belongs to the class of random effects selection models, using latent random variables and common covariates that link together the sub-models.

Huang *et al*. (2009) present the remeasurement method to diagnose random effect model misspecification of the type that leads to biased inference on joint models. The method is derived from the SIMEX method to reveal sensitivity of the target estimator to model assumptions on the random effects. The results are illustrated and compared with application to data for a primary endpoint and a longitudinal process.

Other useful references include Faucett and Thomas (1996), Lavalley and Degrutolla (1996), Faucett *et al*. (1998), Finkelstein and Schoenfeld (1999), Kalbfleisch and Prentice (2002), and Tsiatis and Davidian (2004), Garrett Fitzmaurice *et al*. (2009, Chapter 13).

A further difficulty for making inference on the longitudinal process is that occurrence of the event may induce an informative censoring. It is frequently the case that clinical trials and observational studies involve some missing data. The occurrence of the key event is censored by some competing risk such as disease-related dropout, which could cause non-ignorable missing data. Subjects move away, fail to keep some appointments, or die. Adjustment of inferences about longitudinal measurements to allow for possibly outcome-dependent dropout has been discussed by Wu and Carroll (1988), Hogan and Laird (1997ab), and many other authors. Although the selection models we discussed have been widely applied to both longitudinal and survival studies, another class of models called mixture models appear to be used primarily for studies involved informative dropout. In such cases the mechanisms of the missingness in data need carefully examination. Valid inference requires a framework in which underlying relationships between the event and longitudinal process are explicitly acknowledged. We do not discuss this in detail here.

Our goal is to develop a flexible methodology for handling the categorical longitudinal responses and competing risks time-to-event that characterizes the features of our data – the Nun Study data (Snowdon, 1997) for research on dementia. We start with the random effects dependent selection model formulation of Follman and Wu (1995), extending and adapting it to the Nun Study data. A central feature of our modeling strategy is to postulate a shared random effect γ_i for subject i, and assume that within each subject the two components are conditionally independent given the random effect.

1.2 Multi-state Markov transition model

Progression of chronic diseases is often depicted in terms of distinct pre-clinical and clinical phases from normal. The idea of using a multi-state Markov model to model the transitions among these states and quantify the effects of changes in risk factors is straightforward. In particular, a nonhomogeneous Markov model can be easily applied to model the progression of disease with increasing or decreasing risks by time. Kay (1986) proposed a stochastic process to analyze biomarkers and disease states data in survival studies on cancer. Muenz and Rubinstein (1985) used a Markov chain to model a binary sequence of states and extended the basic model to allow time-dependent covariates. However, there are many circumstances in which estimation of the transition matrix is complicated by the complex relationship among transition probabilities. Craig and Sendi (2002) summarized methods to obtain the maximum likelihood estimate of the transition matrix for discrete-time Markov chains and used the bootstrap method to construct confidence intervals for functions of the transition matrix such as expected survival.

Based on the transitional modeling (Agresti, 2002), Salazar (2004 and 2007) proposed his approach featured in modeling longitudinal categorical responses as a multi-state system where series of categorical outcomes are expressed in terms of states. The onset and progression of these outcomes are modeled as transitions among the states.

For presentation purpose, we assume a finite stochastic system that consists of three transient states and two competing absorbing states. This corresponds to the five progression stages in the study of dementia (Tyas *et al*., 2007). According to Salazar *et al*. (2007), a multinomial logit parameterization could be applied to link the transition probabilities with the fix and random effects.

$$log\left(\frac{P_{sv}(\theta_{sv}|X,\gamma)}{P_{s1}(\theta_{s1}|X,\gamma)}\right) = \alpha_v + X'\beta_v + \xi_v^s + W'\gamma, \quad v = 2,3,4,5$$

The model formulation in terms of logit functions allows us to find a closed form expression for each transition probability and hence to derive the marginal likelihood function based on the conditional distribution of the longitudinal response vector y_i. The likelihood function for the ith subject of his model is

$$L_i(\boldsymbol{\Theta}|X) = \int L_i(\boldsymbol{\Theta}|X,\gamma) \times h(\gamma)d\gamma \approx \int \prod_{i=1}^{n_i} P_{y_{il-1},y_{il}}(\boldsymbol{\Theta}|X,\gamma) \times h(\gamma)d\gamma$$

The overall likelihood function can be obtained by evaluating the product of $L_i(\boldsymbol{\Theta}|X)$ defined by the trajectory of subject i. We will discuss Salazar's modeling approach in more detail in the following chapter. Yu $et\ al.$ (2009) suggested to extend this model to account for the possible dependency between the baseline information and the random effects, and showed improvement in parameter estimation.

1.3 Parameter estimation

Assuming the random effect is normally distributed, the resultant marginal likelihood needs to be evaluated numerically in order to produce parameter estimates. Salazar $et\ al.$ (2007) compared three commonly used techniques for approximating the type of integrations: Laplace approximation (Gao, 2004; Skrondal and Rabe-Hesketh, 2004), Gauss-Hermite quadrature technique (Hedeker and Gibbons, 1994; Skrondal and Rabe-Hesketh, 2004), and importance sampling method (Salazar, 2004). Each of the method is tested using different distributional assumptions for the random effect during the

simulation study. The Gauss quadrature method is recommended in terms of less bias and better confidence interval coverage under all distributional forms of the random effect as well as its computational simplicity.

In numerical analysis, Gauss–Hermite quadrature is an extension of Gaussian quadrature method for approximating the value of integrals from $-\infty$ to ∞ of the kind: $\int_{-\infty}^{\infty} e^{-x^2} f(x)dx$. Like all the other forms of Gaussian quadrature, it solves integrals in a numerical way by approximating the integral with summation using a series of optimal points and weights. In univariate case, the log likelihood function is written as follows

$$logL \approx \log\{\textstyle\sum_q w_q\, e^{b_q^2}\kappa(b_q)\, h(b_q)\}$$

Here w_q and b_q are the corresponding Gaussian weights and abscissas (quadrature points), and $f(\cdot)$ is the probability density function of the random effects term. In multivariate case, the approximation is analogous in the sense that each single quadrature point is replaced with a multi-dimensional vector of quadrature points (Hedeker and Gibbons, 1994). However, the computation can be heavily intensive since the terms in the summation increase exponentially as the dimension of random effects grows. Agresti (2002) proposed to use an adaptive version of Gauss-Hermite quadrature that requires less optimal points and therefore works more efficiently than the ordinary rule. Laplace's method is also deemed to be useful and computationally efficient to construct asymptotic approximations in high dimensional settings.

1.4 Motivations

The research on multi-state Markov transition model is motivated by the nature of the longitudinal data from the Nun Study (Snowdon, 1997), and similar information on the BRAiNS cohort (Salazar, 2004). Here BRAiNS is an acronym for Biologically Resilient Adults in Neurological Studies. Information on the progression of participants at risk for disease is available at unequally spaced points over time during which the conditions of the sisters are assessed and they may transition forward and backward among certain non-absorbing states until diagnosed with the dementia (for instance, Alzheimer's Disease). These transient cognitive states are defined as Intact Cognition, Mild Cognitive Impairment, and Global Impairment in previous work (Salazar, 2004; Tyas *et al.*, 2007). The criteria to classify a nun in a particular transient cognitive state are given below:

Intact Cognition: The patient passes all cognitive and Activities of Daily Living tests.

Mild Cognitive Impairment: The patient passes the Delayed Word Recall, Mini-Mental State Exam, and Activities of Daily Living tests but fails one or more of the other three cognitive tests (Boston Naming, Verbal Fluency, and Constructional Praxis).

Global Impairment: The patient passes the Delayed Word Recall but fails the Mini-Mental State Exam, Activities of Daily Living test, and one or more of the other three cognitive tests (Boston Naming, Verbal Fluency, and Constructional Praxis) without meeting criteria for dementia.

The cognitive test battery is part of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD). The general structure of the Nun Study data is presented in Table 1.1.

Table 1.1 General structures of the Nun Study Data

Subject	Observed State	Cov 1	...	Cov p	Residual Survival	Cov 1	...	Cov q	Shared Random Effect
1	y_{11}	x_{111}	...	x_{1p1}	t_1	z_{11}	...	z_{1q}	γ_1
	$\vdots$	$\vdots$	...	$\vdots$			...		
	y_{1n_1}	x_{11n_1}	...	x_{1pn_1}			...		
2	y_{21}	x_{211}	...	x_{2p1}	t_2	z_{21}	...	z_{2q}	γ_2
	$\vdots$	$\vdots$	...	$\vdots$			...		
	y_{2n_2}	x_{21n_2}	...	x_{2pn_2}			...		
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
i	y_{i1}	x_{i11}	...	x_{ip1}	t_i	z_{i1}	...	z_{iq}	γ_i
	$\vdots$	$\vdots$	...	$\vdots$			...		
	y_{in_i}	x_{i1n_i}	...	x_{ipn_i}			...		
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	$\vdots$
m	y_{m1}	x_{m11}	...	x_{mp1}	t_m	z_{m1}	...	z_{mq}	γ_m
	$\vdots$	$\vdots$	...	$\vdots$			...		
	y_{mn_m}	x_{m1n_m}	...	x_{mpn_m}			...		

In most longitudinal studies on progression to disease when the target population is elderly subjects, death is one of the competing risks. Our analyses on the Nun Study are based on data from the eleven successive examinations, which consists of 672 participants aged 75+ when enrolled in the study. Among the final analytic sample of 461 subjects, 74 (16%) survived without dementia, 162 (35%) developed dementia, and 225 (49%) died before converting to dementia. In order to identify risk factors associated with transitions and thus to determine the probability that a nun with given risk factors will contract dementia before dying, most authors of earlier literature handle death as a competing absorbing state to dementia in the Markov process (Salazar, 2004; Salazar *et al.*, 2007). In contrast with Salazar's model, we propose to model the transition probability with a four state Markov chain, same transient states (Intact Cognition, M.C.I., and G.I.) but dementia being the only absorbing state. We consider incorporating

information on the actual residual survival times from death of the subjects into the stochastic system. Such an approach could allow different risk factors for dementia and death thus to improve the regression estimation since the model likelihood components are built up separately.

Considerable literature can be found that focuses on constructing extended likelihood functions to accommodate missing data that are non-ignorable or informative drop-out (Follman and Wu, 1995; Ten Have *et al.*, 1998 and 2000; Gao, 2004; Vonesh *et al.*, 2006; Li *et al.*, 2007; Shen and Gao, 2007). A popular approach in this respect is to define the shared random effects, given which the two likelihood components the follow-up response and the drop-out response are assumed to be conditionally independent (Ten Have *et al.*, 1998). Recall the random effects dependent selection model formulation of Follman and Wu (1995) that we discussed in Section 1.1, the joint distribution of the follow-up response Y_i and the censored event time D_i for the ith subject can be expressed as

$$f(y_i, d_i) = \int_{b_i} f(y_i|b_i)f(d_i|b_i)dF(b_i)$$

Here the random effects b_i's are assumed to have some prior distribution function form of $F(\cdot)$. This approach was adopted by Yu *et al.* (2009) for the purpose of extending the model likelihood to account for the baseline information. Similarly, we could base the analyses on this model formulation which makes it possible for us to incorporate the residual survival time of the subjects.

One major assumption we made about the residual survival is that the distribution follows a parametric family, Weibull in particular. This raises questions on the validity of inference in the case when the assumption gets violated. It is of interest to investigate in detail how the distributional assumption of the survival would actually affect the parameter estimates in the Markov chain. As a preliminary look at the model assumption, we can compute the estimated cumulative survival curves by Kaplan-Meier estimation method and check the fit statistics. We present a simulation study to further explore the impact of distributional assumption of the survival being violated in terms of estimating bias and MSE. What if the survival times of the subjects come from other common survival distributions, for instance, Log-normal. The influence brought by different sample sizes will also be discussed.

The absorption statistics are of particular interest in a multi-state Markov model. Consider an arbitrary finite nonstationary absorbing Markov chain with state space $U = \{1, 2, \ldots, u\}$. Define T to be the set of transient states and $R = T^c$ the set of absorbing states. Let $P^{(m,m+k)}$ denote the k-step transition matrix with m being the initial starting time of the chain, and $P^{(m,m+k)} = P^{(m,m+1)} \times \ldots \times P^{(m+k-1,m+k)}$, the product of k one-step transition matrices. In the homogeneous cases, the stationary condition holds we end up having $P^{(m,m+k)} = P^k$. If there are r absorbing states and t transient states (so in our case $r=1$ and $t=3$), the one-step transition matrix will have the following canonical form

$$P^{(1)} = \begin{bmatrix} Q_{t \times t} & R_{t \times r} \\ 0_{r \times t} & I_{r \times r} \end{bmatrix}$$

For a homogeneous Markov chain, the fundamental matrix $M = (I - Q)^{-1}$ is well-defined and its elements can be calculated from the converging series $M = I + Q + Q^2 + \cdots$.

However, in situations where the model involves time dependent risk factors such as age, the transition probabilities among states vary with time and the underlying transition probability matrix is no longer homogeneous. The corresponding fundamental matrix of the chain is replaced with an infinite matrix series whose convergence status requires a closer examination before the absorption statistics can be properly calculated; while the survival component confounding in the chain complicates the problem regarding formulation and computation for both the point and interval estimates.

1.5 Outline of the dissertation

The remainder of this dissertation is organized as follows:

In Chapter two we proposed our approach to the problem that incorporate a residual survival from death to Salazar's multi-state Markov model (2007). To do so, we treat the survival from death as a continuous variable rather than defining death as a competing absorbing state to dementia. We assume that within each subject the survival component and the Markov process are linked by a shared latent random effect, and moreover, these two pieces are conditionally independent given the random effect and their corresponding predictor variables. Then a closed-form expression for the individual and thus overall conditional marginal likelihood function is derived, which we can evaluate numerically to produce the maximum likelihood estimates for the unknown parameters. Later in the

chapter we present the results of the simulation studies that design to show how the model's ability to accurately estimate the parameters can be affected by the distributional form of the survival term. Finally, we illustrate the results with an application to the Nun Study data. We discuss our findings and further provide the results by adding the missing portion of the baseline responses previously suggested by Yu *et al*. (2009) as comparison.

In Chapter three we consider an extended nonhomogeneous Markov transition model. We focus on addressing the problem by accommodating the residual life time of the subjects confounding in the nonhomogeneous chain. The convergence status of the chain is examined and the formulation of the absorption statistics (1) probability of developing dementia before death p, and (2) relative risk of absorption between the two competing events dementia and death $RR = \frac{p}{1-p}$, are derived. Then we propose using the Delta method to estimate the variance terms to construct confidence intervals for p and the odds ratio $OR = \left(\frac{p_1}{1-p_1}\right) / \left(\frac{p_2}{1-p_2}\right)$. Since the technique is based on the assumption of the asymptotic normal sampling distribution, we carefully check for normality with simulated samples (set to have 10,000 iterations). The results are illustrated with the Nun study data in detail.

Finally in Chapter four we summarize the most relevant findings, state the advantages and disadvantages of our methodology, and provide the areas for future research.

Chapter 2 : A Markov transition model with death as a competing event

2.1 Introduction

In clinical trials and observational studies, it is common that the occurrence of the key event is censored by some competing risk such as disease-related dropout, which could cause non-ignorable missing data. More specifically, in most longitudinal studies on progression to a certain disease when the target population is elderly subjects, death is one of the competing risks. In Nun study among the total of 461 subjects – the final analytic sample for parameter estimating, almost half ($n = 225$) died before converting to dementia. Several existing approaches have been developed in joint analysis of the longitudinal measurements and competing risks time-to-event data (Elashoff 2007; Xu and Zeger 2001). However, few involve categorical responses that characterize our data.

Salazar (2007) proposed a suitable approach to the problem by defining a multi-state Markov chain to model the progression of dementia in which death was treated as a competing absorbing state to dementia. A possible alternative is to model the survival from death as continuous variable. We consider incorporating a Weibull survival to Salazar's Markov model assuming a shared random effect. A closed-form expression for the conditional marginal likelihood function is derived. The model stability to the violation of the assumption on distributional form of survival is tested in simulation studies.

The chapter is organized as follows: in Section 2.2 we construct the model likelihood function; in Section 2.3 we present the results of the simulation studies; in Section 2.4 we apply the model to the Nun study data; and in Section 2.5 we summarize our findings.

2.2 Model and estimation

2.2.1 Salazar's multi-state Markov model

Suppose there are m subjects in the study. For subject i let $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{in_i})$ denote the random vector representing the observed cognitive states for subject i at n_i different ordered discrete occasions, where $i = 1, 2, \dots, m$. We assume the Markov property holds, that is, the conditional distribution of $f(y_{ik}|y_{i1}, \dots, y_{ik-1})$ is identical to the conditional distribution of $f(y_{ik}|y_{ik-1})$ for $k = 1, 2, \dots, n_i$. Then conditioned on Y_{i1}, the joint distribution of the random vector $\boldsymbol{Y}_i$ can be written as

$$f(\boldsymbol{y}_i|y_{i1}) = f(y_{i2}, y_{i3}, \dots, y_{in_i}|y_{i1}) = f(y_{i2}|y_{i1})\, f(y_{i3}|y_{i2}) \dots f(y_{in_i}|y_{in_{i-1}})$$

Here the subscript y_{ik} refers to the state occupied by the ith subject at kth occasion. In order to simplify the notation, we can use $P_{y_{ik}, y_{ik-1}} = f(y_{ik}|y_{ik-1})$ to denote the one step transition probability from state y_{ik-1} to state y_{ik}. So for instance, if $y_{ik-1} = s$ and $y_{ik} = v$, then P_{sv} represents the probability of transition for subject i from state s to state v during the $k-1$th and kth visits. Throughout, we use upper-case letters to represent random variables and lower-case letters for their realizations; dependence on covariates is usually suppressed for notational clarity.

In the example to be discussed later – the nun study data, the status of a participant at each visit was recorded as being one of the states: intact cognition, mild cognitive impairments (M.C.I.), global impairments (G.I.), or dementia (Tyas et. al., 2007). The participants were followed during the study period until death occurred. The conditional distribution of the status of an individual participant at an arbitrary examination given her status at previous examinations was assumed to have the Markov property, i.e., that status at the examination depended on only the most recent previous examination and was independent of status at other previous examinations. Following Salazar et. al. (2007), a multi-state Markov chain was used to model transitions from one state to another, in which intact cognition, mild cognitive impairments, and global impairments were considered transient states, whereas dementia and death were absorbing states as shown in Figure 2.1.
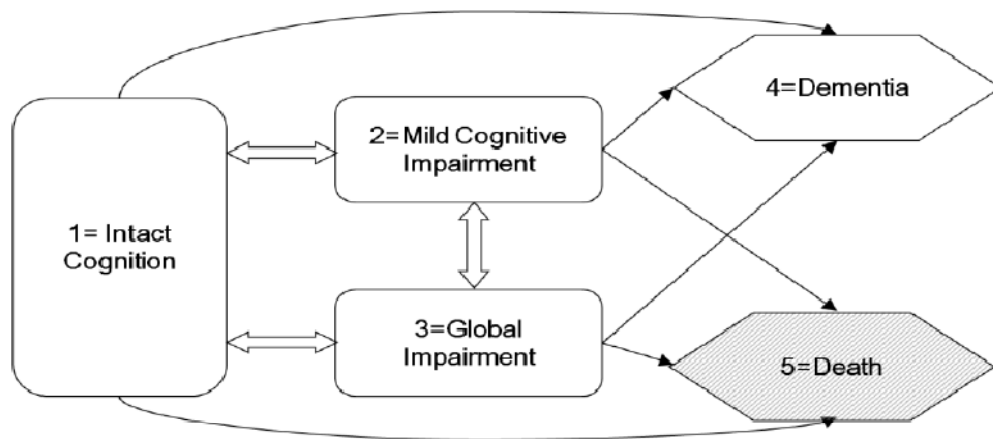
Figure 2.1 Possible cognitive transitions between three transient states (1) intact cognition (2) M.C.I. (3) G.I. and two absorbing states (4) dementia (5) death

Thus the one-step transition probability matrix could be presented in the form of

$$
\begin{bmatrix}
P_{11}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{12}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{13}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{14}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{15}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) \\
P_{21}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{22}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{23}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{24}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{25}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) \\
P_{31}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{32}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{33}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{34}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) & P_{35}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

According to Salazar et. al. (2007), a multinomial logit parameterization could be applied to link these transition probabilities with the fixed and random effects.

$$
\log\left(\frac{P_{sv}(\boldsymbol{\theta}_{sv}|X,\boldsymbol{\gamma})}{P_{s1}(\boldsymbol{\theta}_{s1}|X,\boldsymbol{\gamma})}\right) = \alpha_v + X'\boldsymbol{\beta}_v + \xi_v^s + W'\boldsymbol{\gamma}, \quad v = 2,3,4,5.
$$

Here $\boldsymbol{\Theta}$ represents the set of all the unknown parameters, $\boldsymbol{\alpha}$ is the vector of intercepts, $\boldsymbol{\beta}$ is the vector of unknown fixed effects for covariates X, and $\boldsymbol{\xi}$ is the set of unknown fixed effects for the prior state. Also, $\boldsymbol{\gamma}$ is the vector of unobserved random effects associated with subject i. The formulation of Salazar's model in terms of logit functions allows us to find the closed expression for each transition probability as follows

$$
P_{sv}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma}) =
\begin{cases}
\dfrac{1}{1 + \sum_{h=2}^{4} \exp\left(\alpha_h + X'\boldsymbol{\beta}_h + \xi_h^s + W'\boldsymbol{\gamma}\right)} & for\ v = 1 \\[3mm]
\dfrac{\exp\left(\alpha_v + X'\boldsymbol{\beta}_v + \xi_v^s + W'\boldsymbol{\gamma}\right)}{1 + \sum_{h=2}^{4} \exp\left(\alpha_h + X'\boldsymbol{\beta}_h + \xi_h^s + W'\boldsymbol{\gamma}\right)} & for\ v > 1
\end{cases}
$$

Therefore, based on the conditional distribution of $f\left(y_{i2}, y_{i3}, \dots, y_{in_i} \mid y_{i1}\right)$ the marginal likelihood function for the ith subject is

$$
L_i(\boldsymbol{\Theta}|X) = \int_{\Omega} \prod_{l=2}^{n_i} \times \prod_{\substack{s=1\cdots3, \\ v=1\cdots5}} \left(P_{sv}(\boldsymbol{\Theta}|X,\boldsymbol{\gamma})\right)^{\delta_{y_{il-1},s}\delta_{y_{il},v}} h(\boldsymbol{\gamma})d\boldsymbol{\gamma} \tag{2.1}
$$

with Ω denoting the support for the distribution of the random vector $\boldsymbol{\gamma}$. The probability density function for $\boldsymbol{\gamma}$ is $h(\cdot)$. Here $\delta_{y_{il-1},s}$ and $\delta_{y_{il},v}$ are indicator functions valued at 1 if $y_{il-1} = s$ and $y_{il} = v$, and 0 otherwise. The overall likelihood function can be obtained by evaluating the product of (2.1) across the subjects under study. However, this approach may lead to biased or inconsistent estimates since the likelihood is based on the conditional distribution instead of the full distribution in which the baseline information is ignored.

2.2.2 Model improvement with Weibull survival

In Salazar's model death is modeled as the competing absorbing state to dementia. A possible alternative approach is to incorporate information on the actual survival times from death of the subjects into the stochastic system.

Xu and Zeger (2001) proposed a latent variable model to model the relationship between time-to-event data, longitudinal response, and covariates, in which covariates could only affect the longitudinal response through its influence on an assumed latent process. Elashoff et. al. (2006) suggested joint modeling of the repeated measures and competing risk failure time data by using latent random variables and common covariates to link the sub-models.

However, in our case the data involves multinomial responses and the parameterization of a polychotomous logit under a discrete time Markov framework complicates the problem. We hypothesize that the survival times of the subjects come from certain parametric distribution which shares the same random effects used in Markov transition

model. Additionally these two pieces are conditionally independent given the random effects and their corresponding predictor variables.

In contrast with Salazar's model, we are modeling the transition probability with a four state Markov chain, same transient states but dementia being the only absorbing state. The one-step transition probability matrix now becomes

$$\begin{bmatrix} P_{11}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{12}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{13}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{14}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) \\ P_{21}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{22}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{23}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{24}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) \\ P_{31}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{32}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{33}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) & P_{34}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Each transition probability P_{sv} could be postulated in the form of

$$P_{sv}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}) = \begin{cases} \dfrac{1}{1 + \sum_{h=2}^{4} \exp(\alpha_h + \boldsymbol{X}'\boldsymbol{\beta}_h + \xi_h^s + \boldsymbol{W}'\boldsymbol{\gamma})} & for\ v = 1 \\[4mm] \dfrac{\exp(\alpha_v + \boldsymbol{X}'\boldsymbol{\beta}_v + \xi_v^s + \boldsymbol{W}'\boldsymbol{\gamma})}{1 + \sum_{h=2}^{4} \exp(\alpha_h + \boldsymbol{X}'\boldsymbol{\beta}_h + \xi_h^s + \boldsymbol{W}'\boldsymbol{\gamma})} & for\ v > 1 \end{cases}$$

We further assume the survival time $\boldsymbol{S} \sim Weibull(r > 0, \boldsymbol{\mu})$, where $\boldsymbol{\mu} = e^{\eta_0 + \boldsymbol{Z}'\boldsymbol{\eta} + \boldsymbol{W}'\boldsymbol{\gamma}}$. The probability of the ith subject failing from the competing risk of death is

$$\pi(S_i = t_i|\boldsymbol{\Theta},\boldsymbol{\gamma}) = \left[r e^{\eta_0 + \boldsymbol{\eta}'Z_i + w_i'\gamma_i} t_i^{r-1} exp\left(-e^{\eta_0 + \boldsymbol{\eta}'Z_i + w_i'\gamma_i} t_i^r\right) \right]^{\tau_i} \left[exp\left(-e^{\eta_0 + \boldsymbol{\eta}'Z_i + w_i'\gamma_i} t_i^r\right) \right]^{1-\tau_i}$$

Here τ_i is some indicator function valued at 1 if the ith subject died at time t_i and 0 otherwise. $\boldsymbol{\Theta}$ is the parameter vector associated with both the transition probability and the probability of death. For each subject under study, the conditional marginal likelihood function for the ith subject can be rewritten as

$$L_i(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{Z}) = \int \prod_{l=2}^{n_i} \times \prod_{\substack{s=1\cdots3, \\ v=1\cdots4}} \left(P_{sv}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma})\right)^{\delta_{y_{il-1},s}\delta_{y_{il},v}} \times \pi_{S_i}(\boldsymbol{\Theta}|\boldsymbol{Z},\boldsymbol{\gamma})h(\boldsymbol{\gamma})d\boldsymbol{\gamma}.$$

2.2.3 Parameter estimation

Assuming that the random effect is distributed as a $N(0, \sigma^2)$, the resultant log likelihood can be maximized using the Gauss-Hermite quadrature method combined with the Newton-Raphson method to numerically evaluate the derivatives and produce the parameter estimates. The estimates of the standard errors are computed by Fisher's information method.

2.3 Simulations

The main purpose of the simulation study is to examine the sensitivity of the MLEs to the violations of the Weibull model assumption on the survival time. We want to quantify how the distributional form for the survival term affects the model estimates associated with the fixed effects. In addition, the model's ability to accurately estimate the unknown parameters is of interest. To answer these questions we look at two aspects: (i) the bias of the MLEs to the true parameters and (ii) the mean squared errors of the MLEs.

The simulation was set to have 300 iterations, with each containing 200 or 500 subjects. Each subject has up to ten follow-up waves starting from a baseline state of intact cognition. We considered four cases:

1. Total of 200 subjects generated with prior distribution of survival being Weibull
2. Total of 500 subjects generated with prior distribution of survival being Weibull
3. Total of 200 subjects generated with prior distribution of survival being Lognormal
4. Total of 500 subjects generated with prior distribution of survival being Lognormal

In the cases when the true prior distribution of survival being lognormal (case 3 and 4), the probability density functions were plotted along with the simulated Weibull distribution for comparison. As shown in Figure 2.2, the red line representing the lognormal densities from which the survival times of the simulation populations in case 3 and 4 were generated; and the blue line indicating the simulated Weibull distribution that we fit in our model. The four plots correspond with the four possible combinations of the values from the two model covariates. The relative location and diversity of the two curves differ by the values of the covariates.

Thus two sets of comparison could be made to explore: first, the effects of varying the sample size, and second, the effects of violating the original model assumption on the distributional form of survival term with a possible alternative.

In both situations, the transition probabilities were dependent on three covariates age, prior state (intact cognition or M.C.I. or G.I.), and the presence/absence of an apolipoprotein E-4 allele (APOE4). The covariates entered in the survival model were age at entry and the APOE4 status of the subject. All the simulations were done using the IML procedure in SAS system. The results are presented in Table 2.1.

Figure 2.2 Probability densities of the prior distribution of survival (True vs. Simulated)
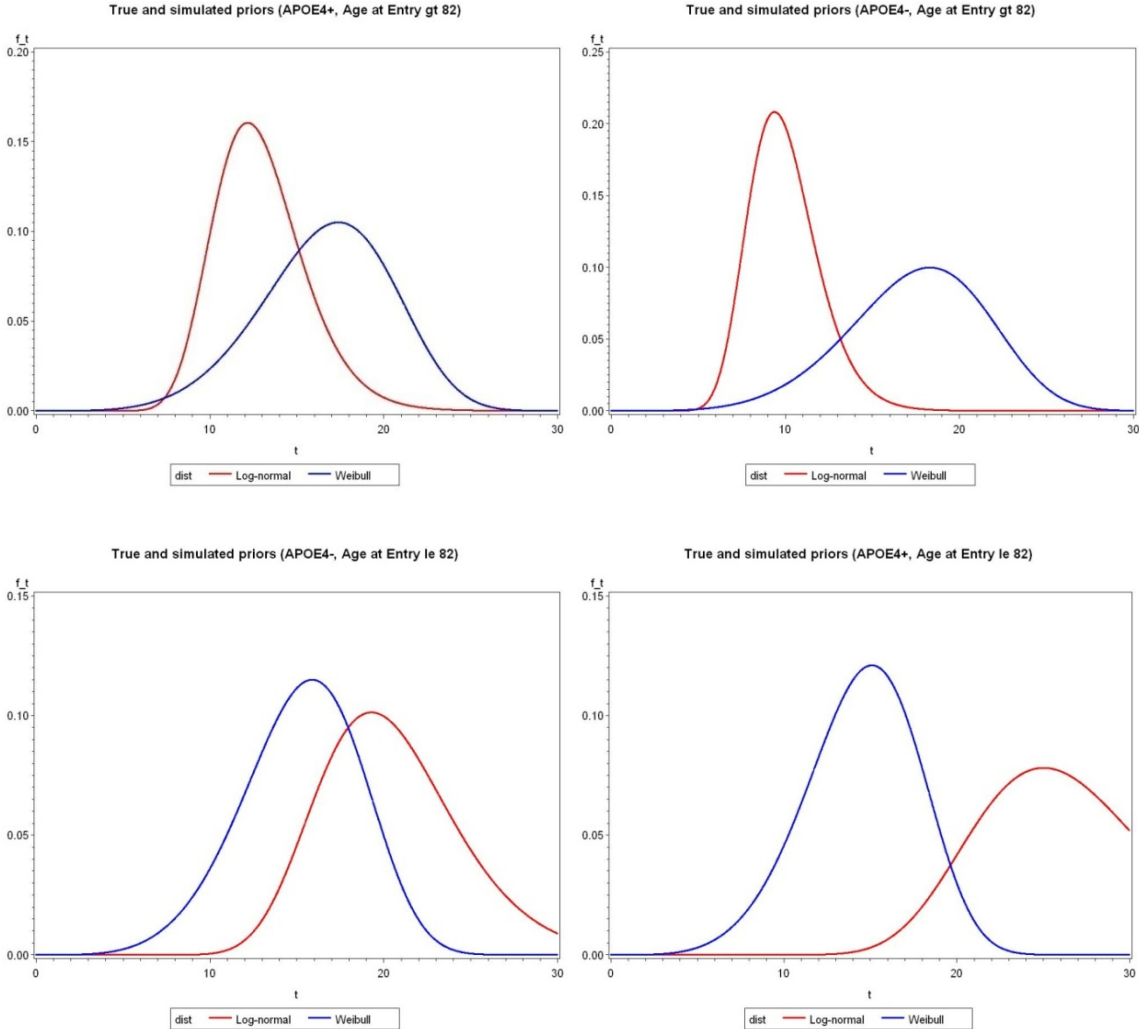
Table 2.1 Bias and Mean squared error of the model parameters

(Base state: 1 = Intact Cognition)

Risk Factors	State	True Parameter	Weibull survival (200 subjects)		Weibull survival (500 subjects)	
			Bias	MSE	Bias	MSE
Age	2	0.07	-0.0035	0.00046	-0.0017	0.00024
	3	0.19	0.0039	0.00097	0.0025	0.00040
	4	0.21	0.0713	0.01030	0.0593	0.00520
APOE4	2	0.27	-0.0854	0.07311	-0.0636	0.03143
	3	0.39	-0.1208	0.14037	-0.1092	0.04940
	4	1.62	0.0221	0.35017	-0.0065	0.09960
Prior state:						
Intact Cognition	2	0.58	0.1389	0.30218	0.0884	0.08680
	3	-3.12	0.0547	0.09054	0.0142	0.03484
	4	-4.15	-0.2528	0.55932	-0.1517	0.21298
Mild Cognitive	2	1.79	0.1139	0.30504	0.0683	0.08110
Impairment	3	-2.43	0.0142	0.11078	0.0033	0.04011
	4	-2.26	-0.1115	0.40921	-0.0953	0.14037

Risk Factors	State	True Parameter	Lognormal survival (200 subjects)		Lognormal survival (500 subjects)	
			Bias	MSE	Bias	MSE
Age	2	0.07	0.0185	0.00070	0.0226	0.00064
	3	0.19	0.0259	0.00150	0.0334	0.00147
	4	0.21	0.0997	0.01372	0.0939	0.01032
APOE4	2	0.27	-0.2442	0.10083	-0.1760	0.04622
	3	0.39	-0.2733	0.16793	-0.2665	0.09779
	4	1.62	-0.0983	0.20157	-0.1794	0.10875
Prior state:						
Intact Cognition	2	0.58	-0.2302	0.27467	-0.2802	0.14252
	3	-3.12	-0.2799	0.16309	-0.2801	0.11369
	4	-4.15	-0.9934	0.75993	-0.4577	0.82227
Mild Cognitive	2	1.79	0.0222	0.20501	0.0036	0.06647
Impairment	3	-2.43	-0.0134	0.08151	0.0208	0.03255
	4	-2.26	-0.0513	0.22739	-0.0255	0.09625

In general, increasing the sample size would improve the estimates in terms of reducing bias and MSE when the Weibull is the true distribution as we assumed. In the case when the prior distribution of survival being lognormal instead, only MSE was influenced by increasing the sample size from 200 to 500, while bias did not change much. Moreover, the results indicate that the maximum likelihood estimates are not sensitive to violations of the assumed Weibull model in the case when the lognormal is the true distribution.

2.4 Application to the Nun Study

The Nun Study began enrollment in 1991. The data consists of a cohort of 672 members of the School Sisters of Notre Dame born before 1917 and living in retirement communities in the Midwestern, eastern, and southern United States. The subjects were recruited in phases and received periodic cognitive assessments with brain donation at death. Analyses were based on data from the eleven successive examinations. A total of 211 subjects were excluded from the study due to: missing examinations, missing APOE data, or presence of dementia at baseline visit. The final analytic sample consisted of 461 participants, of which 74 survived without dementia, 162 developed dementia, and 225 died before converting to dementia. The transitions among the cognitive states are summarized in Table 2.2 below.

Table 2.2 Number of transitions in the Nun Study

	Current Visit			
Prior Visit	Intact Cognition	M.C.I.	G.I.	Dementia
Intact Cognition	593 (69.9%)	197 (23.2%)	54 (6.3%)	5 (0.6%)
M.C.I.	177 (16.2%)	697 (63.8%)	136 (12.5%)	82 (7.5%)
G.I.	16 (5.1%)	39 (12.4%)	184 (58.6%)	75 (23.9%)
Dementia	0 (0%)	0 (0%)	0 (0%)	81 (100%)

The covariates of interest were age, education level, APOE4 status, and prior state. For simplicity purposes education was not included in the model in our simulations; but was considered here since it is a well-known risk factor and found to be significantly associated with dementia in previous studies. The covariates entered in the survival model were age at entry and APOE4 status. As shown in Figure 2.3 below, subjects were subgrouped based on their APOE4 status and age at entry, and thus four Weibull probability plots were created as a preliminary look at the model assumption. The estimated cumulative distribution function was computed by Kaplan-Meier estimation in the Lifereg procedure in SAS. The straight line represents the maximum likelihood fit, with the simultaneous parametric confidence bands on each side. The values of the censored observations are plotted along the top of each graph in red. The plots indicate that the assumed Weibull model fits the data reasonably well although not perfect since skewness arises in the tail of the distribution for some of the groups.

Figure 2.3 Weibull probability plots of the survival time in the Nun Study
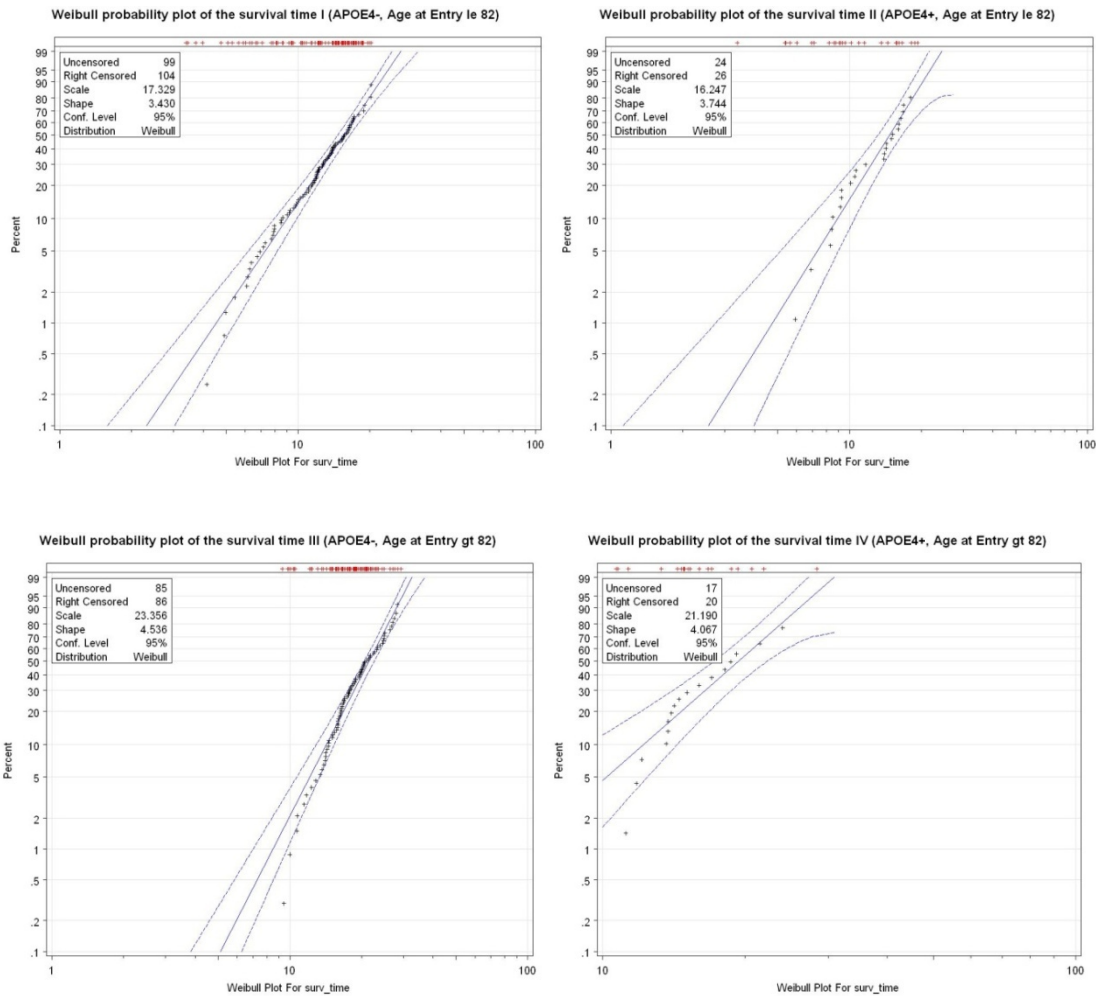
Similarly assuming the survival time of the subjects follows a lognormal distribution, the data was fitted and tested with the same covariates. It suggests that Weibull model is a better fit to the data. The lognormal probability plots for the four subgroups are illustrated in Figure 2.4.

Figure 2.4 Lognormal probability plots of the survival time in the Nun Study
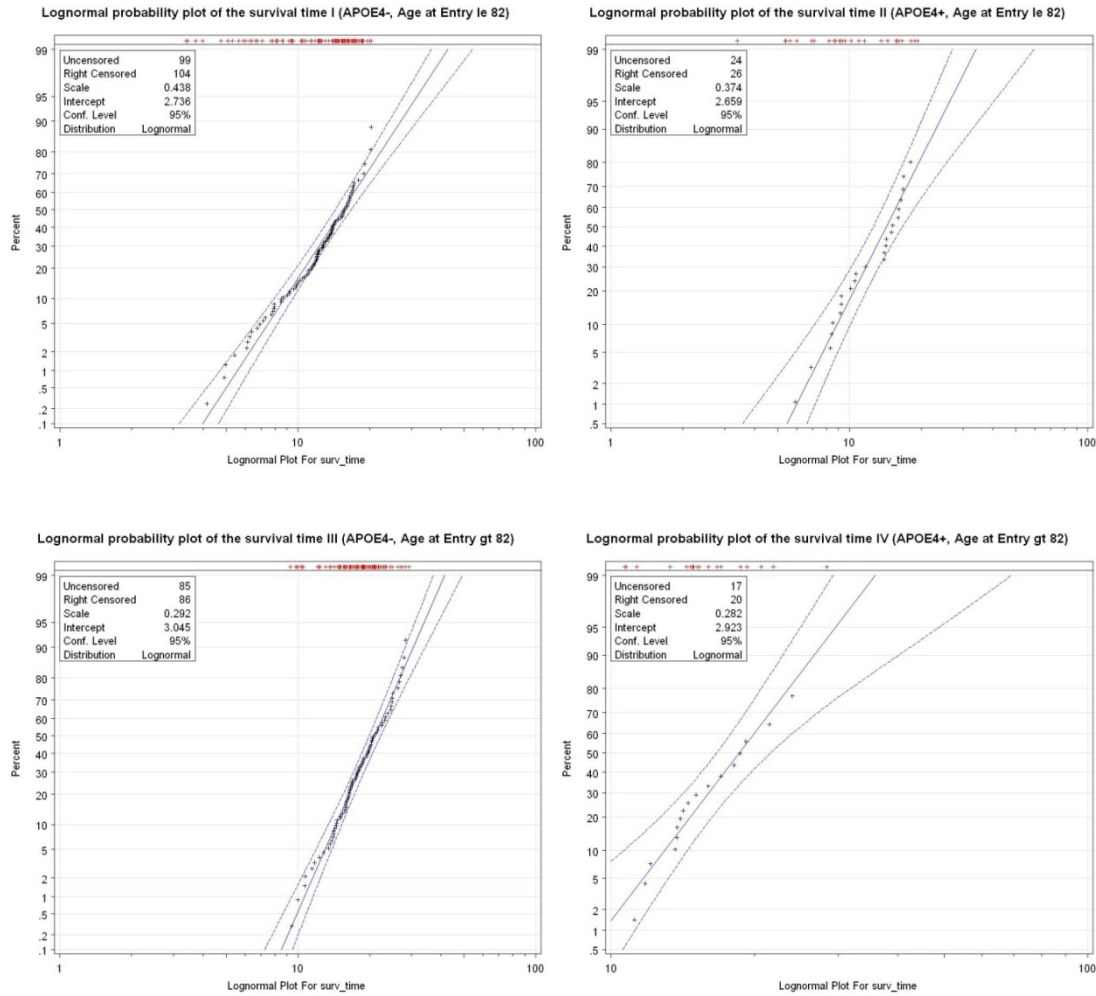
Table 2.3 lists the parameter estimates for the transition probabilities. First, as expected Age and APOE4 are significant predictors of a transition to M.C.I., G.I., and Dementia as opposed to a transition to cognitively normal because all the coefficients associated with Age and APOE4 are significant. The odds ratios OR_{Age}=(1.10, 1.19, 1.18) and OR_{APOE4}=(2.31, 3.64, 4.10) are significantly different from one. Second, remaining cognitively intact favors the highly educated. Compared to those with more than 16 years of education, subjects with 12 years or less have significant odds ratios for transitions to

M.C.I. (OR=4.55) and G.I. (OR=4.77); Similarly, the corresponding odds ratios to M.C.I. (OR=1.59) and G.I. (OR=1.64) are significant for those with exactly 16 years of education. These results are consistent with those from previous studies. Moreover, Age and APOE4 are both significant predictors for survival time but education is not. The coefficient associated with Age is negative indicating that an elderly age at entry could be "protective" for subjects from the competing risk of death. That is, the likelihood of dementia before death increases with age.

Table 2.3 Maximum likelihood estimates (SE) of model parameters in the Nun Study

(Base state: 1 = Intact Cognition)

Risk Factors	State	Estimates	Std. Error	Risk Factors	State	Estimates	Std. Error
Markov chain				Prior states:			
Age	2	0.092*	(0.016)	Intact Cognition	2	-1.232*	(0.334)
	3	0.172*	(0.020)		3	-3.834*	(0.326)
	4	0.169*	(0.023)		4	-5.344*	(0.545)
APOE4	2	0.838*	(0.232)	Mild Cognitive	2	0.670*	(0.327)
	3	1.292*	(0.263)	Impairment	3	-2.375*	(0.306)
	4	1.412*	(0.297)		4	-1.997*	(0.327)
Education:				Weibull survival			
< 16 years	2	1.515*	(0.348)	Age at Entry	-	-1.523*	(0.183)
vs. > 16 years	3	1.562*	(0.391)	APOE4	-	0.447*	(0.220)
	4	1.403*	(0.436)	Rate	-	4.613*	(0.272)
16 years	2	0.465*	(0.158)				
vs. > 16 years	3	0.497*	(0.194)	Sigma	-	0.871*	(0.100)
	4	0.372	(0.235)				

* Significant at P < 0.05

29

2.5 Conclusion and discussion

Considerable literature has focused on characterizing the relationship between longitudinal response process and time-to-event data. In contrast, relatively little research has been done to accommodate multinomial responses, with even fewer relying on a polychotomous logit parameterization under a discrete-time Markov chain.

As an improvement to Salazar's multi-state Markov model, we fit a Weibull model to the survival from death and correlate it with the Markov transition model by defining a shared random effect. The simulation study showed model stability in terms of violations of the distributional assumption on survival time. More specifically, the maximum likelihood estimates are not sensitive to violations of the assumed Weibull model if, in fact, a lognormal model should be used instead.

The application to the Nun study data found that Age and APOE4 are significant predictors of a transition to impaired states and Dementia as opposed to a transition to cognitively normal because all the coefficients associated with Age and APOE4 are significant and positive. Remaining cognitively intact favors the highly educated (> 16 years education) which also agrees with the results from the previous models. Age and APOE4 are both significant predictors for survival time. Age at entry is "protective" for subjects from the competing risk of death since older subjects are more likely to become demented before death.

Yu et. al. (2009) incorporated the missing portion of the baseline likelihood into the follow-up likelihood by assuming the two share the same random effect. The complete marginal likelihood function for the ith subject with baseline can be written as

$$L_i(\boldsymbol{\Theta}, \boldsymbol{\Phi}|X) = \int \prod_{l=2}^{n_i} \times \prod_{\substack{s=1\cdots3, \\ v=1\cdots5}} \left(P_{sv}(\boldsymbol{\Theta}|X, \boldsymbol{\gamma})\right)^{\delta_{y_{il-1},s}\delta_{y_{il},v}} \times \pi_{y_{i1}}(\boldsymbol{\Phi}|X_B, \boldsymbol{\gamma})h(\boldsymbol{\gamma})d\boldsymbol{\gamma}$$

Here $\boldsymbol{\Phi}$ is the set of parameters associated with the baseline response components. The probability of the baseline state $\pi_{y_{i1}}(\boldsymbol{\Phi}|X_B, \boldsymbol{\gamma})$ was similarly modeled by using multinomial logistic regression as for the one-step transition probability $P_{sv}(\boldsymbol{\Theta}|X, \boldsymbol{\gamma})$ in the follow-up likelihood. It will also be interesting to combine this approach with our model to find a complete likelihood function that accommodates all the three pieces baseline, follow-up, and survival.

The results from the Nun study data are presented in Table 2.4 below. Note that in this application 81 subjects who were diagnosed with dementia at the baseline visit entered to help estimate the baseline effects. Those subjects were dropped from our previous model without the baseline likelihood component.

Table 2.4 Maximum likelihood estimates (SE) of model parameters with baseline

Risk Factors	State	Estimates	Std. Error	Risk Factors	State	Estimates	Std. Error
Markov chain				*Baseline*			
Age	2	0.118*	(0.017)	*Age*	2	0.121*	(0.034)
	3	0.199*	(0.020)		3	0.269*	(0.040)
	4	0.196*	(0.023)		4	0.270*	(0.037)
APOE4	2	1.078*	(0.263)	*APOE4*	2	0.642	(0.367)
	3	1.536*	(0.291)		3	1.400*	(0.447)
	4	1.649*	(0.322)		4	1.703*	(0.400)
Education:				*Education:*			
< 16 years	2	1.829*	(0.388)	*< 16 years*	2	2.215*	(0.623)
vs. > 16 years	3	1.873*	(0.427)	*vs. > 16 years*	3	2.588*	(0.706)
	4	1.713*	(0.468)		4	3.247*	(0.653)
16 years	2	0.590*	(0.177)	*16 years*	2	0.660*	(0.287)
vs. > 16 years	3	0.618*	(0.209)	*vs. > 16 years*	3	0.458	(0.399)
	4	0.493*	(0.247)		4	0.693*	(0.352)
Prior states:				Weibull survival			
Intact Cognition	2	-0.837*	(0.339)	Age at Entry	-	-1.701*	(0.198)
	3	-3.437*	(0.330)	APOE4	-	0.552*	(0.248)
	4	-4.947*	(0.548)	Rate	-	5.081*	(0.279)
Mild Cognitive	2	0.701*	(0.335)				
Impairment	3	-2.346*	(0.314)	Sigma	-	1.212*	(0.085)
	4	-1.968*	(0.335)				

* Significant at P < 0.05

We can see that although the new model produced similar values of the parameter estimates as our previous model, the magnitude of odds ratios are larger for all the risk factors Age, APOE4 status, education level and prior state under the new model with baseline. For example, keeping other covariates constant, the odds ratio of having APOE4 present for transitions from intact cognition to M.C.I. is 2.31, to G.I. is 3.64 and to dementia is 4.10 under the previous model. In comparison, the corresponding odds ratios are 2.94, 4.65 and 5.20 under the current model with baseline likelihood

component. The results are similar for the other risk factors in the model. The comparison

of fit statistics presented in Table 2.5 also suggests that the inclusion of the baseline

component might help make up for those information potentially missing from our

previous model and improve the parameter estimates.

Table 2.5 Comparison of fit statistics in the Nun Study data

Fit Statistics	Model wo Baseline - Weibull	Model w Baseline - Weibull	Model w Baseline - Lognormal
-2 Log Likelihood	6855.2	6814.3	7027.9
AIC	6937.2	6896.3	7109.9
AICC	6938.7	6897.8	7111.4
BIC	7113.3	7072.4	7286.0

Future extensions of the model may include considering the random-effects models, in

which less strict assumptions about the association between the two outcomes Y_1 and Y_2

are required. The general idea is to define separate but correlated latent variables b_1 and

b_2 for Y_1 and Y_2, and let $f(b_1, b_2)$ denote the joint density (often bivariate normal).

Then by assuming conditional independence of Y_1 and Y_2 given (b_1, b_2), the joint

density of (Y_1, Y_2) can be obtained from

$$f(y_1, y_2) = \iint f(y_1|b_1)f(y_2|b_2)f(b_1, b_2)db_1 db_2.$$

A more flexible setting is to consider using a proportional hazard model for the residual

survival time of the subject depending on fixed effects z_i and random effect b_{2i}

$$h_i(t|b_{2i}) = h_0(t)\exp(z_i'\beta + \alpha'b_{2i})$$

with piecewise-constant baseline hazard (step-functions): $h_0(t) = \sum_{q=1}^{Q} \xi_q I(v_{q-1} < t \le v_q)$ where $0 = v_0 < v_1 < \cdots < v_Q$ denotes a split of the time scale.

Chapter 3 : Extended nonhomogeneous Markov transition model

3.1 Introduction

In the study of chronic diseases like Alzheimer's, it is commonly the case that the investigators are particularly interested in the probability of disease onset before dying given a set of risk factors such as age, education, and genetic status. The purpose of this chapter is to continue the study for an extended nonhomogeneous Markov transition model that involves time dependent risk factors as well as the survival component. In that case, the underlying transition probability matrix is no longer stationary. The convergence status of the chain needs further examination before the absorption statistics can be computed.

The remainder of the chapter is organized as follows: Section 3.2 introduces notations for defining common absorption statistics; investigates the convergence status of the fundamental matrix series and derives the formulas to compute the probability of dementia before death. Section 3.3 illustrates the use of the Delta method to construct confidence intervals for the transition probabilities to dementia and the odds ratios. In Section 3.4 the results are applied to the Nun Study. Section 3.5 compares the results under a simplified model with the risk indicator variable that combines the effects of the original risk factors APOE4 and Education. Conclusions are summarized in Section 3.6.

3.2 Nonhomogeneous Markov chain

3.2.1 Notation and definition

A homogenous Markov chain has lots of nice properties and attributes. We can argue that in the homogeneous case the absorption of transient states is guaranteed and the absorption statistics can be calculated explicitly. However, the chain considered here involves time dependent risk factors such as age, in which case the transition probabilities among states vary with time and thus the underlying transition probability matrix is no longer stationary. In this section, we continue to investigate for the convergence status and the statistics characterizing transitions and absorptions among states for a nonhomogeneous chain.

Now consider an arbitrary finite absorbing nonstationary Markov chain with state space $U = \{1,2,\dots,u\}$. Define T to be the set of transient states and $R = T^c$ the set of absorbing states. Let $P^{(m,m+k)}$ denote the k-step transition matrix with m being the initial starting time of the chain, and $P^{(m,m+k)} = P^{(m,m+1)} \times \dots \times P^{(m+k-1,m+k)}$, the product of k one-step transition matrices. $P^{(m,m+k)}$ can also be expressed in the following canonical form:

$$P^{(m,m+k)} = \begin{bmatrix} Q^{(m,m+1)} & R^{(m,m+1)} \\ 0 & I \end{bmatrix} \times \dots \times \begin{bmatrix} Q^{(m+k-1,m+k)} & R^{(m+k-1,m+k)} \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} \prod_{l=1}^{k} Q^{(m+l-1,m+l)} & R^{**} \\ 0 & I \end{bmatrix}$$

here $R^{**} = R^{(m,m+1)} + Q^{(m,m+1)}R^{(m+1,m+2)} + \dots + Q^{(m,m+k-1)}R^{(m+k-1,m+k)}$. Q and R are the substochastic matrices describing transitions among the transient states and transitions from the transient states to the absorbing states, respectively.

Suppose the chain starts from transient state i at time m. Note that when the stationary condition holds we have $\boldsymbol{P}^{(m,m+k)} = \boldsymbol{P}^k$ since transitions in the chain no longer depend on time.

The following subsections will focus on addressing the problem by accommodating the residual survival time of the subject confounding in the nonhomogeneous chain. The formulation of the absorption statistics and the construction of their confidence intervals are discussed in detail.

3.2.2 Probability of dementia before death

Suppose that a certain subject has initially started the process from state i, $i \in \boldsymbol{T}$. Let Y_n be the state that the process visits at time n. Let $W = \min_n(Y_n = j)$, $j \in \boldsymbol{R}$, which is the time it takes for the process to enter an absorbing state. W is only observed at each visit time. Further define S to be the residual life time of this subject at the time he/she enrolled in the study. S is continuous.

Recall that in the previous chapter we hypothesize that the residual survival times of the subjects come from known parametric distribution sharing the same random effects used in Markov chain, i.e. here $f_S(\cdot \,|\boldsymbol{\theta}, \boldsymbol{\gamma})$ and $F_S(\cdot \,|\boldsymbol{\theta}, \boldsymbol{\gamma})$ being the conditional Weibull probability density function and cumulative distribution function, respectively.

Let m denote the initial starting time. For each fixed t, the probability that absorption occurs after time t can thus be derived. $\forall\, i \in \boldsymbol{T}$, we have

$$P(W \le t|\boldsymbol{\Theta},\boldsymbol{\gamma}) = P(W \le t^*|\boldsymbol{\Theta},\boldsymbol{\gamma}) = \sum_{v=1}^{t^*} P(W = v|\boldsymbol{\Theta},\boldsymbol{\gamma})$$

$$= \left[\boldsymbol{R}^{(m,m+1)} + \sum_{v=1}^{t^*-1} \left(\prod_{l=1}^{v} \boldsymbol{Q}^{(m+l-1,m+l)} \right) \boldsymbol{R}^{(m+v,m+1+v)} \right](i)$$

$$= \left[\sum_{v=0}^{t^*-1} \boldsymbol{Q}^{(m,m+v)} \boldsymbol{R}^{(m+v,m+1+v)} \right](i)$$

where $t^* = \lfloor t \rfloor = \max\{x \epsilon \mathbb{Z} | x \le t\}$.

$$P(W = t^*|\boldsymbol{\Theta},\boldsymbol{\gamma}) = [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i)$$

So, $\forall\, i \in \boldsymbol{T}$,

$$P(W < S|\boldsymbol{\Theta},\boldsymbol{\gamma}) = \sum_{t^*=1}^{\infty} P(W = t^*, S > t^*|\boldsymbol{\Theta},\boldsymbol{\gamma}) = \sum_{t^*=1}^{\infty} P(W = t^*|\boldsymbol{\Theta},\boldsymbol{\gamma})\, P(S > t^*|\boldsymbol{\Theta},\boldsymbol{\gamma})$$

$$= \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot \int_{t^*}^{\infty} f_S(t|\boldsymbol{\Theta},\boldsymbol{\gamma})dt$$

$$= \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\Theta},\boldsymbol{\gamma})]$$

Yu et al. (2009) showed that under certain conditions the norm of the substochastic matrix $\|\boldsymbol{Q}^{(m,n)}\|$ converges to zero as $n \to \infty$. Following the property of cumulative distribution functions $0 \le F_S(t^*|\boldsymbol{\Theta},\boldsymbol{\gamma}) \le 1$, we have

$$\forall\, i \in \boldsymbol{T}, \quad P(W < S|\boldsymbol{\Theta},\boldsymbol{\gamma}) \le \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i)$$

The probabilities of dementia before death converge with time.

We have $\qquad\qquad P(W > S|\boldsymbol{\Theta},\boldsymbol{\gamma}) = 1 - P(W < S|\boldsymbol{\Theta},\boldsymbol{\gamma})$

Therefore, by taking the integral over the whole support for the distribution of the random vector $\boldsymbol{\gamma}$, we have

$$P(W < S|\boldsymbol{\Theta}) = \int \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\Theta}, \boldsymbol{\gamma})] h(\boldsymbol{\gamma}) \, d\boldsymbol{\gamma}$$

where $h(\cdot)$ denotes the probability density function for $\boldsymbol{\gamma}$.

Given the random effects, the relative risk of absorption between the two competing events can be derived by taking the ratio

$$RR = \frac{P(W < S|\boldsymbol{\Theta})}{P(W > S|\boldsymbol{\Theta})}$$

$$= \frac{\int \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\Theta}, \boldsymbol{\gamma})] \, h(\boldsymbol{\gamma}) d\boldsymbol{\gamma}}{1 - \int \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)} \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\Theta}, \boldsymbol{\gamma})] \, h(\boldsymbol{\gamma}) d\boldsymbol{\gamma}}$$

The resultant integral can be solved numerically using the Gauss-Hermite quadrature method as discussed in Section 1.3.

3.3 Construction of confidence intervals

Our primary research interest in this study is to estimate the confidence intervals associated with the probabilities and odds ratios of developing dementia before death. One such approach is using the Delta method to estimate the corresponding standard errors and construct the confidence intervals based on the assumption of the normal sampling distribution.

It is necessary for us to check the normality of the estimated p and OR in the Nun Study before applying the method. In the case when the distribution of p (or OR) is skewed from normal, we will also look at possible transformations, such as logarithm of the statistic, for a better interval estimate.

3.3.1 Checking for normality

As shown in the diagram below, the simulation is performed in the following steps:

<u>Step1</u>. The maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ as well as the associated covariance matrix $cov(\widehat{\boldsymbol{\beta}})$ can be derived from our model as discussed in the previous chapter;

<u>Step2</u>. Each individual vector of $\widetilde{\boldsymbol{\beta}}_i$ is generated from the multivariate normal distribution $MVN(\widehat{\boldsymbol{\beta}}, cov(\widehat{\boldsymbol{\beta}}))$;

<u>Step3</u>. The transition probability matrices $\widetilde{\boldsymbol{P}}_i^{(m,m+t)}(\boldsymbol{X}|\widetilde{\boldsymbol{\beta}})$ and the cumulative survival distribution functions $F_S(t|\widetilde{\boldsymbol{\beta}})$ are then estimated;

<u>Step4</u>. The probability of developing dementia before death can be calculated from the submatrices $\widetilde{\boldsymbol{Q}}_i^{(m,m+t-1)}(\boldsymbol{X}|\widetilde{\boldsymbol{\beta}})$, $\widetilde{\boldsymbol{R}}_i^{(m+t-1,m+t)}(\boldsymbol{X}|\widetilde{\boldsymbol{\beta}})$, and $F_S(t|\widetilde{\boldsymbol{\beta}})$;

<u>Step5</u>. The p_is and OR_is as well as their functional transformations i.e., $arcsin\sqrt{p_i}$ and $\ln(OR_i)$ are computed;

<u>Step6</u>. Repeat the above steps for n times;

Step7. Standard univariate procedures are conducted to evaluate and compare the normality of these statistics p_i, OR_i, $arcsin\sqrt{p_i}$, and $\ln(OR_i)$.

Figure 3.1 Diagram of the data generation procedure
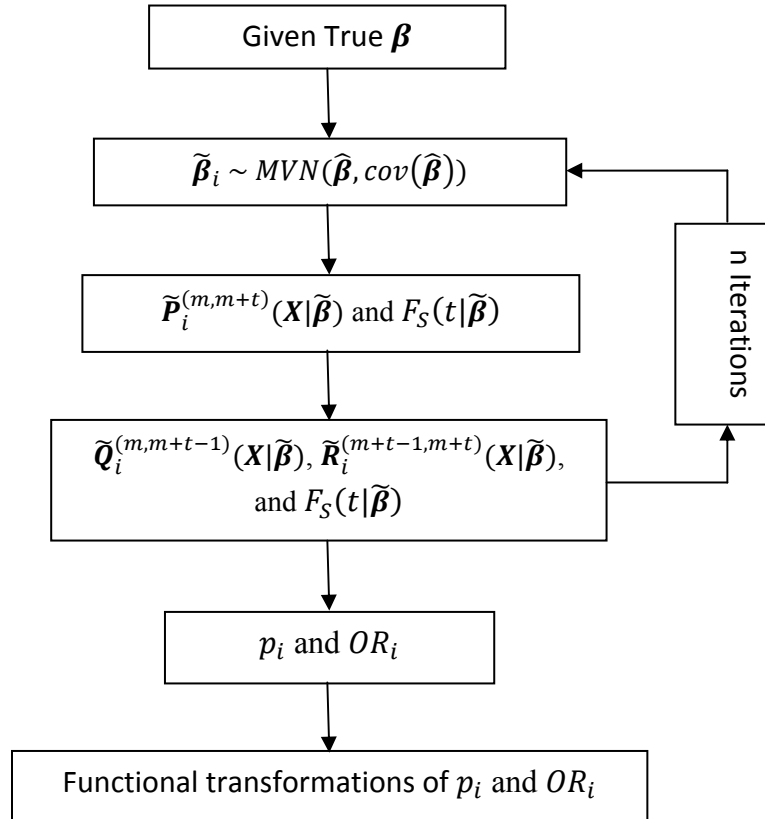
For illustration purposes, we pick a typical nun with baseline age 80, apoe4 positive, less than 16 years of education, and initial state being intact cognition. (The results for initial state being M.C.I. or G.I. are quite similar.) The number of iterations is 10,000.

We are looking at the sampling distribution of $p = \Pr(dementia\ before\ death)$, the odds ratio $OR = \dfrac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$, and the log or other type of transformations of them. To compute the odds ratio, p_2 is taken from a nun with same baseline age, education level, and initial state, but apoe4 negative. The main output is presented as follows:

41

Figure 3.2 Checking for normality of *p* in the Nun Study
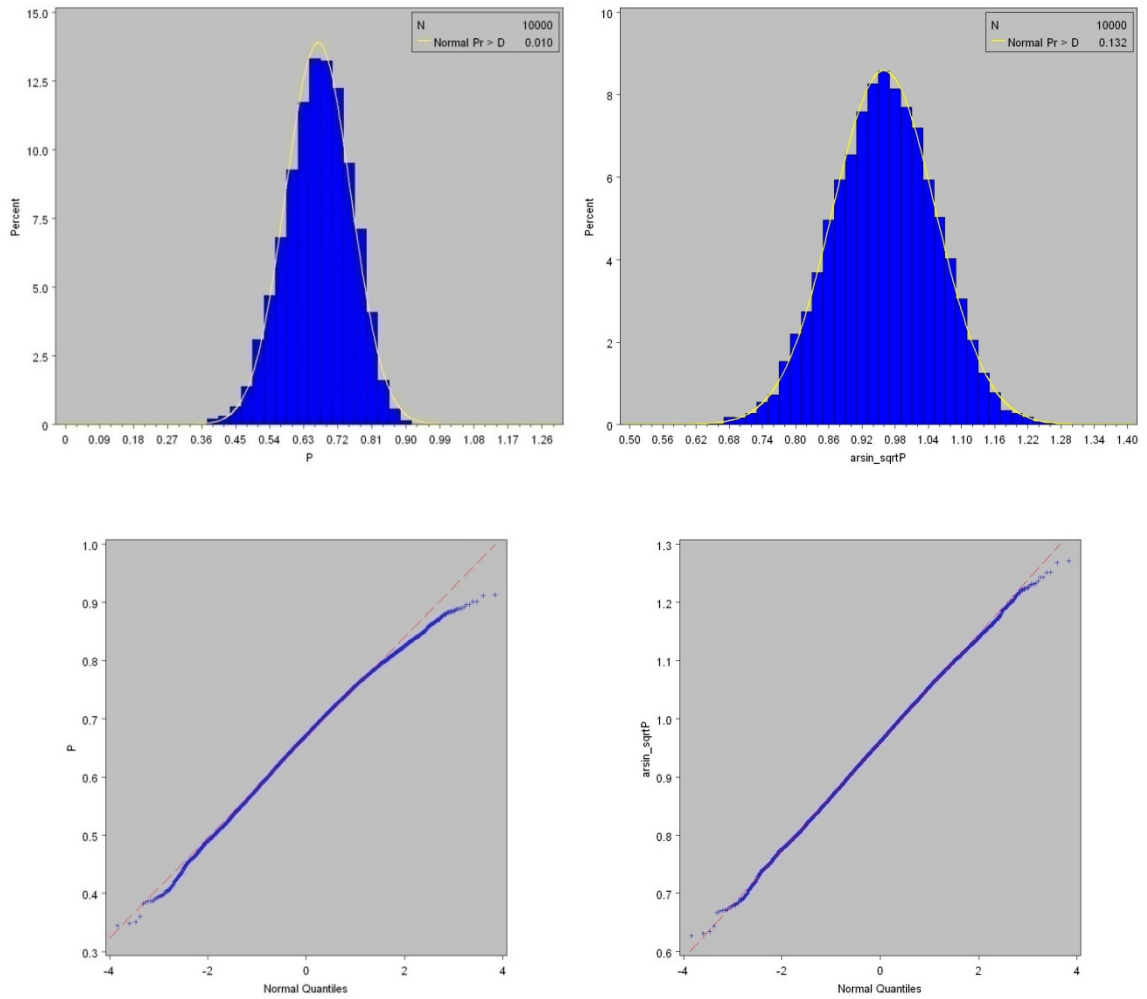

```
             Goodness-of-Fit Tests for Normal Distribution

Test                    ----Statistic-----     ------p Value------

Kolmogorov-Smirnov      D       0.00791052     Pr > D         0.132
Cramer-von Mises        W-Sq    0.15989975     Pr > W-Sq      0.019
Anderson-Darling        A-Sq    1.16387869     Pr > A-Sq      <0.005
```

Figure 3.3 Checking for normality of OR in the Nun Study
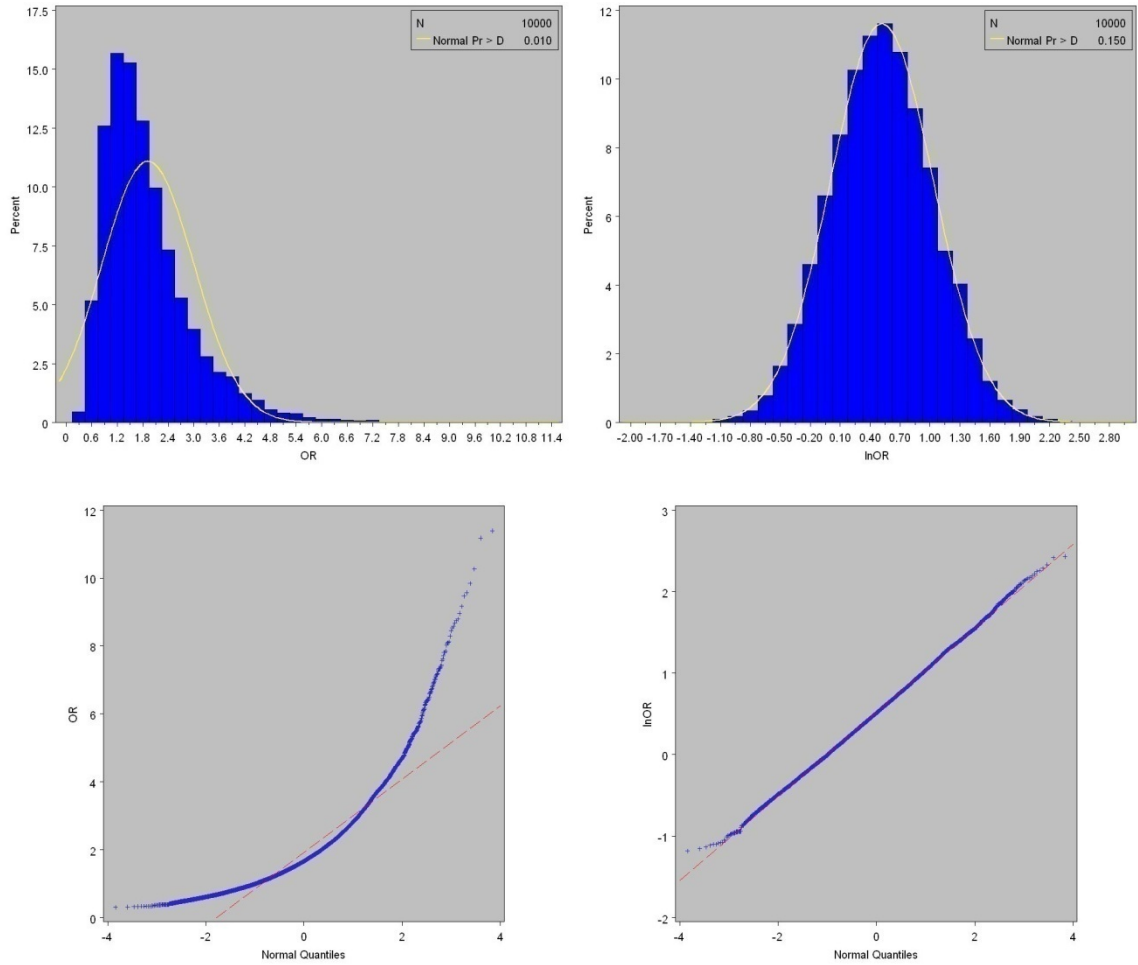

```
           Goodness-of-Fit Tests for Normal Distribution

    Test                    ----Statistic-----    ------p Value------

    Kolmogorov-Smirnov    D       0.00686499    Pr > D        >0.150
    Cramer-von Mises      W-Sq    0.12037271    Pr > W-Sq      0.063
    Anderson-Darling      A-Sq    0.86690532    Pr > A-Sq      0.026
```

The results suggest that direct estimates of p, ln (p), $\frac{p}{1-p}$, and $ln\left(\frac{p}{1-p}\right)$ are not asymptotic

normal while in comparison $arcsin\sqrt{p}$, the inverse trigonometric function of the square

root of p, is more likely to be normally distributed. Moreover, the sampling distribution

for ln (OR) is also normal, the hypothesis testing of normality is not rejected.

Consequently, based on asymptotic normal assumption, we would thus look at $arcsin\sqrt{p}$

and $\ln(OR)$ to construct the confidence interval for p and OR.

3.3.2 Applying the Delta method

(i) s.e. for $arcsin\sqrt{p}$ using Delta method

Let $p = \Pr(developing\ dementia\ before\ death)$, we have shown that

$$P(W < S|\boldsymbol{\Theta},\boldsymbol{\gamma}) = \sum_{t^*=1}^{\infty} [\boldsymbol{Q}^{(m,m+t^*-1)}\ \boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\Theta},\boldsymbol{\gamma})] \qquad (3.1)$$

By Delta method, if $g(f(p)) = arcsin\sqrt{p}$ then

$$var(arcsin\sqrt{p}) \approx [g'(f(p)) \cdot f'(p)]^2 var(\hat{p}) \qquad (3.2)$$

$$= \left(\frac{1}{\sqrt{1-p^2}} \cdot \frac{1}{2\sqrt{p}}\right)^2 var(\hat{p}) \qquad (3.3)$$

$$= \left(\frac{1}{2\sqrt{p-p^3}}\right)^2_{\hat{p}} var(\hat{p}) \qquad (3.4)$$

$$= \left[\frac{1}{4(p-p^3)}\right]_{\hat{p}} var(\hat{p}) \qquad (3.5)$$

Here we can estimate $var(\hat{p})$ by a second round of Delta method. i.e. we need to start

with Eq. (3.1) and the covariance matrix of the estimated parameter vector $\hat{\boldsymbol{\beta}}$.

(ii) s.e. for $\ln(OR)$ using Delta method

Let $p_1 =$ Probability of dementia in group 1, and

$p_2 =$ Probability of dementia in group 2, we have $OR = \frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}$.

44

By Delta method,

$$var\left(\frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}\right) = \left(\frac{1}{\frac{p_2}{1-p_2}}\right)^2 \frac{1}{(1-p_1)^4} var(p_1) + \left(\frac{p_1}{1-p_1}\right)^2 \left(-\frac{1}{p_2^2}\right)^2 var(p_2)$$

$$-2\left(\frac{1-p_2}{p_2^3}\right)\left[\frac{p_1}{(1-p_1)^3}\right] cov(p_1,p_2) \tag{3.6}$$

Then if we apply Delta method again with respect to the logarithm, we have

$$var\left(ln\left(\frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}\right)\right) = \left(\frac{\frac{p_2}{1-p_2}}{\frac{p_1}{1-p_1}}\right)^2 \cdot var\left(\frac{\frac{p_1}{1-p_1}}{\frac{p_2}{1-p_2}}\right) \tag{3.7}$$

$$\approx \left[\frac{1}{p_1^2(1-p_1)^2}\right]_{\hat{p}_1,\hat{p}_2} var(\hat{p}_1) + \left[\frac{1}{p_2^2(1-p_2)^2}\right]_{\hat{p}_1,\hat{p}_2} var(\hat{p}_2)$$

$$-2\left[\frac{1}{p_1 p_2(1-p_1)(1-p_2)}\right]_{\hat{p}_1,\hat{p}_2} cov(\hat{p}_1,\hat{p}_2) \tag{3.8}$$

3.3.3 Formulating the covariance matrix

Let $\boldsymbol{F}^{(m,m+t^*)} = (1 - F_S(t^*|\boldsymbol{\Theta},\boldsymbol{\gamma})) \otimes \boldsymbol{I}_3$, then

$$P(W < S|\boldsymbol{\Theta},\boldsymbol{\gamma}) = \sum_{t^*=1}^{\infty}[\boldsymbol{Q}^{(m,m+t^*-1)}\,\boldsymbol{R}^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\Theta},\boldsymbol{\gamma})]$$

$$= \sum_{t^*=1}^{\infty}[\boldsymbol{F}^{(m,m+t^*)}\boldsymbol{Q}^{(m,m+t^*-1)}\,\boldsymbol{R}^{(m+t^*-1,m+t^*)}](i)$$

Define $\boldsymbol{P}^{(m)} = \sum_{t^*=1}^{\infty}[\boldsymbol{F}^{(m,m+t^*)}\boldsymbol{Q}^{(m,m+t^*-1)}\,\boldsymbol{R}^{(m+t^*-1,m+t^*)}] \equiv \boldsymbol{P}^{(m,X)}$, $\boldsymbol{X}$ is the covariate

matrix associated with $\boldsymbol{P}$. For instance, given $\boldsymbol{X} = (age\ 80, APOE4+, EDUC <$

$16\ years)$, the three entries of $\boldsymbol{P}^{(m)}$ will be the probability of developing dementia

before death for a certain subject with baseline age 80, apoe4 positive, less than 16 years

of education starting from the initial state of intact cognition, M.C.I., and G.I.,

respectively.

Now let $\boldsymbol{P}^{*(m)} = \begin{pmatrix} \boldsymbol{P}^{(m,X)} \\ \boldsymbol{P}^{(m,X^*)} \end{pmatrix}$. $\boldsymbol{P}^{*(m)}$ is a 6×1 vector, which contains the probabilities

from two different sets of covariates. We have

$$cov\left(\boldsymbol{P}^{*(m)}\right) \approx \left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{*(m)}\right] V(\boldsymbol{\beta}) \left[\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{*(m)}\right]^t$$

$$= \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m,X)} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m,X^*)} \end{bmatrix} V(\boldsymbol{\beta}) \begin{bmatrix} \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m,X)} \\ \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m,X^*)} \end{bmatrix}^t .$$

In the following context we will address how to derive $\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m,X)}$ through matrix

calculation. $\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m,X^*)}$ can be solved analogously by simply substituting $\boldsymbol{X}$ with $\boldsymbol{X}^*$.

Assume the process consists of 3 transient states (normal, MCI, GI) and 1 absorbing state

(dementia),

- denote $\frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{Q}_i)$ by $\boldsymbol{A}(i)$, $\boldsymbol{A}(i)$ is $9 \times l$ (given l is the length of $\boldsymbol{\beta}$)

- denote $\frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{R}_i)$ by $\boldsymbol{B}(i)$, $\boldsymbol{B}(i)$ is $3 \times l$

- denote $\frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{F}_i)$ by $\boldsymbol{C}(i)$, $\boldsymbol{C}(i)$ is also $9 \times l$

Hence, we have

$$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{P}^{(m)} = \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_1 \boldsymbol{R}_1 + \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_2 \boldsymbol{Q}_1 \boldsymbol{R}_2 + \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_3 \boldsymbol{Q}_1 \boldsymbol{Q}_2 \boldsymbol{R}_3 + \cdots + \frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_k \prod_{i=1}^{k-1} \boldsymbol{Q}_i \boldsymbol{R}_k$$

Introduce the vector transformation as follows:

$$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_1 \boldsymbol{R}_1 = (\boldsymbol{R}_1^t \otimes \boldsymbol{I}_3) \frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{F}_1) + \boldsymbol{F}_1 \frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{R}_1) = (\boldsymbol{R}_1^t \otimes \boldsymbol{I}_3) \boldsymbol{C}(1) + \boldsymbol{F}_1 \boldsymbol{B}(1)$$

which gives a $3 \times p$ matrix;

$$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_2 \boldsymbol{Q}_1 \boldsymbol{R}_2 = (\boldsymbol{R}_2^t \otimes \boldsymbol{I}_3) \frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{F}_2 \boldsymbol{Q}_1) + (\boldsymbol{F}_2 \boldsymbol{Q}_1) \frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{R}_2)$$

$$= (\boldsymbol{R}_2^t \otimes \boldsymbol{I}_3)[(\boldsymbol{Q}_1^t \otimes \boldsymbol{I}_3)\boldsymbol{C}(2) + (\boldsymbol{I}_3 \otimes \boldsymbol{F}_2)\boldsymbol{A}(1)] + (\boldsymbol{F}_2 \boldsymbol{Q}_1)\boldsymbol{B}(2)$$

also $3 \times p$;

Similarly, we can get

$$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F}_3 \boldsymbol{Q}_1 \boldsymbol{Q}_2 \boldsymbol{R}_3 = (\boldsymbol{R}_3^t \otimes \boldsymbol{I}_3) \frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{F}_3 \boldsymbol{Q}_1 \boldsymbol{Q}_2) + (\boldsymbol{F}_3 \boldsymbol{Q}_1 \boldsymbol{Q}_2) \frac{\partial}{\partial \boldsymbol{\beta}} vec(\boldsymbol{R}_3)$$

$$= (\boldsymbol{R}_3^t \otimes \boldsymbol{I}_3) \left[(\boldsymbol{Q}_2^t \otimes \boldsymbol{I}_3) \frac{\partial}{\partial \boldsymbol{\beta}} (\boldsymbol{F}_3 \boldsymbol{Q}_1) + (\boldsymbol{I}_3 \otimes \boldsymbol{F}_3 \boldsymbol{Q}_1)\boldsymbol{A}(2) \right] + (\boldsymbol{F}_3 \boldsymbol{Q}_1 \boldsymbol{Q}_2)\boldsymbol{B}(3)$$

$$= (\boldsymbol{R}_3^t \otimes \boldsymbol{I}_3)[(\boldsymbol{Q}_2^t \otimes \boldsymbol{I}_3)((\boldsymbol{Q}_1^t \otimes \boldsymbol{I}_3)\boldsymbol{C}(3) + (\boldsymbol{I}_3 \otimes \boldsymbol{F}_3)\boldsymbol{A}(1)) + (\boldsymbol{I}_3 \otimes \boldsymbol{F}_3 \boldsymbol{Q}_1)\boldsymbol{A}(2)]$$

$$+ (\boldsymbol{F}_3 \boldsymbol{Q}_1 \boldsymbol{Q}_2)\boldsymbol{B}(3)$$

Continuing in this manner, for general $k > 1$, the partials for a particular element

$\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{F} \prod_{i=1}^{k-1} \boldsymbol{Q}_i \boldsymbol{R}_k$ can be formulated as

$$(\boldsymbol{R}_k^t \otimes \boldsymbol{I}_3) \left[\prod_{i=1}^{k-1} (\boldsymbol{Q}_i^t \otimes \boldsymbol{I}_3)\boldsymbol{C}(k) + \sum_{i=1}^{k-1} \left(\prod_{p=i+1}^{k-1} (\boldsymbol{Q}_p^t \otimes \boldsymbol{I}_3) \right) \left(\boldsymbol{I}_3 \otimes \boldsymbol{F}_k \prod_{q=0}^{i-1} \boldsymbol{Q}_q \right) \boldsymbol{A}(i) \right]$$

$$+ \boldsymbol{F}_k \prod_{i=1}^{k-1} \boldsymbol{Q}_i \boldsymbol{B}(k)$$

here, define $\boldsymbol{Q}_0 = \boldsymbol{I}_3$ and $\prod_{p=k}^{k-1}(\boldsymbol{Q}_p^t \otimes \boldsymbol{I}_3) = \boldsymbol{I}_9$.

Given the specifications of estimated transition matrix $\boldsymbol{P}$ and the cumulative distribution function of the residual survival time F_S, we can decompose the parameter vector $\boldsymbol{\beta}$ into

i) Transition parameters that are invariant across the rows (e.g. intercept, age, education, apoe4)

ii) Transition parameters that vary across the rows (e.g. prior state indicator: prior1 and prior2)

iii) Survival parameters (e.g. intercept, age at entry, apoe4 in the survival part)

Assume again a process with 3 transient states (normal, MCI, GI) and 1 absorbing state

(dementia). The $\boldsymbol{Q}$ matrix can be specified as

$$
\begin{pmatrix}
\dfrac{1}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior1)_h)} & \dfrac{\exp(\boldsymbol{X\beta}_2+\beta(prior1)_2)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior1)_h)} & \dfrac{\exp(\boldsymbol{X\beta}_3+\beta(prior1)_3)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior1)_h)} \\
\dfrac{1}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior2)_h)} & \dfrac{\exp(\boldsymbol{X\beta}_2+\beta(prior2)_2)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior2)_h)} & \dfrac{\exp(\boldsymbol{X\beta}_3+\beta(prior2)_3)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior2)_h)} \\
\dfrac{1}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h)} & \dfrac{\exp(\boldsymbol{X\beta}_2)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h)} & \dfrac{\exp(\boldsymbol{X\beta}_3)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h)}
\end{pmatrix}
$$

$\boldsymbol{R}$ matrix can be specified similarly

$$
\begin{pmatrix}
\dfrac{\exp(\boldsymbol{X\beta}_4+\beta(prior1)_4)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior1)_h)} \\
\dfrac{\exp(\boldsymbol{X\beta}_4+\beta(prior2)_4)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior2)_h)} \\
\dfrac{\exp(\boldsymbol{X\beta}_4)}{1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h)}
\end{pmatrix}
$$

And in $\boldsymbol{F}$ matrix, $F_{ii}=1-F_S=\exp\left(-e^{\boldsymbol{Z\beta}_s}s^r\right)$, for $i=1,2,3$;

$$F_{ij}=0,\ \text{for}\ i\neq j\ and\ i,j=1,2,3.$$

For a $\beta_2\in\boldsymbol{\beta}_2$ which are the β estimates associated with the transition to state 2 (MCI),

the partial of individual element of $\boldsymbol{Q}$ given X is presented

$$
\frac{\partial}{\partial\beta_2}Q_{11}=\frac{-\exp(\boldsymbol{X\beta}_2+\beta(prior1)_2)\times X}{(1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior1)_h))^2}
$$

$$
\frac{\partial}{\partial\beta_2}Q_{21}=\frac{-\exp(\boldsymbol{X\beta}_2+\beta(prior2)_2)\times X}{(1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h+\beta(prior2)_h))^2}
$$

$$
\frac{\partial}{\partial\beta_2}Q_{31}=\frac{-\exp(\boldsymbol{X\beta}_2)\times X}{(1+\sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h))^2}
$$

$$\frac{\partial}{\partial\beta_2}Q_{12} = \frac{(1 + \sum_{h\neq2}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h)) \times \exp(\boldsymbol{X\beta}_2 + \beta(prior1)_2) \times X}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h))^2}$$

$$\frac{\partial}{\partial\beta_2}Q_{22} = \frac{(1 + \sum_{h\neq2}\exp(\boldsymbol{X\beta}_h + \beta(prior2)_h)) \times \exp(\boldsymbol{X\beta}_2 + \beta(prior2)_2) \times X}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior2)_h))^2}$$

$$\frac{\partial}{\partial\beta_2}Q_{32} = \frac{(1 + \sum_{h\neq2}\exp(\boldsymbol{X\beta}_h)) \times \exp(\boldsymbol{X\beta}_2) \times X}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h))^2}$$

$$\frac{\partial}{\partial\beta_2}Q_{13} = \frac{-\exp(\boldsymbol{X\beta}_3 + \beta(prior1)_3) \times \exp(\boldsymbol{X\beta}_2 + \beta(prior1)_2) \times X}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h))^2}$$

$$\frac{\partial}{\partial\beta_2}Q_{23} = \frac{-\exp(\boldsymbol{X\beta}_3 + \beta(prior2)_3) \times \exp(\boldsymbol{X\beta}_2 + \beta(prior2)_2) \times X}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior2)_h))^2}$$

$$\frac{\partial}{\partial\beta_2}Q_{33} = \frac{-\exp(\boldsymbol{X\beta}_3) \times \exp(\boldsymbol{X\beta}_2) \times X}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h))^2}$$

Other partials of the β_hs invariant across rows can be formulated similarly.

For a $\beta(prior1)_2$ the partial of individual element of $\boldsymbol{Q}$ given X is presented

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{11} = \frac{-\exp(\boldsymbol{X\beta}_2 + \beta(prior1)_2)}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h))^2}$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{21} = 0$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{31} = 0$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{12} = \frac{(1 + \sum_{h\neq2}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h)) \times \exp(\boldsymbol{X\beta}_2 + \beta(prior1)_2)}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h))^2}$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{22} = 0$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{32} = 0$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{13} = \frac{-\exp(\boldsymbol{X\beta}_3 + \beta(prior1)_3) \times \exp(\boldsymbol{X\beta}_2 + \beta(prior1)_2)}{(1 + \sum_{h=2}^{4}\exp(\boldsymbol{X\beta}_h + \beta(prior1)_h))^2}$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{23} = 0$$

$$\frac{\partial}{\partial\beta(prior1)_2}Q_{33} = 0$$

Partials of $\beta(prior1)_h$ and $\beta(prior2)_h$, $h = 2,3$ can be formulated similarly.

Partials of βs in i) and ii) with respect to $\boldsymbol{R}$ can be done analogously.

For a $\beta_{s2} \in \boldsymbol{\beta}_s$ which are the β estimates associated with the survival components, the partial of individual element of $\boldsymbol{F}$ given Z and s is presented

$$\frac{\partial}{\partial\beta_{s2}}F_{ii} = -e^{Z\boldsymbol{\beta}_s}s^r \exp\left(-e^{Z\boldsymbol{\beta}_s}s^r\right) \cdot Z \quad for\ i = 1,2,3$$

$$\frac{\partial}{\partial\beta_{s2}}F_{ij} = 0 \quad for\ i \neq j\ and\ i,j = 1,2,3$$

Other partials of the $\boldsymbol{\beta}_s$ can be formulated similarly.

Partials of βs in i) and ii) with respect to $\boldsymbol{F}$ are zeros.

Partials of βs in iii) with respect to $\boldsymbol{Q}$ and $\boldsymbol{R}$ are zeros.

Back to Eq. (3.5), the variance term for $arcsin\sqrt{p}$ can be estimated by selecting the corresponding diagonal element $var(\hat{p})$ from the covariance matrix $cov\left(\boldsymbol{P}^{*(m)}\right)$. Similarly, if we plug in $\boldsymbol{X} = (age\ 80, APOE4+, EDUC < 16\ years)$ and $\boldsymbol{X}^* = (age\ 80, APOE4-, EDUC < 16\ years)$, we have $var(\hat{p}_1)$, $var(\hat{p}_2)$, and $cov(\hat{p}_1, \hat{p}_2)$, and can then get the variance estimate for $\ln(OR_{APOE4})$ of the two groups by Eq. (3.8). We have shown in the previous section that the statistics $arcsin\sqrt{p}$ and $\ln(OR)$ in our study satisfy the asymptotic normal assumption. Therefore, we can construct confidence intervals for p and OR based on these statistics.

3.4 Application to the Nun Study Data

In this section, we present the analysis of data in the Nun Study described in Section 2.4.
The methodology of computing the probability that a subject develops dementia before
death and thus the relative risk is now applied. The final analytic sample for parameter
estimating consisted of 461 non-demented participants at baseline, of which almost half
(n=225) died before converting to dementia. The 81 subjects diagnosed with dementia at
their baseline visit were excluded.

We let the transition probabilities depend on four covariates: age, the prior state (intact
cognition or M.C.I. or G.I.), the APOE4 status (presence/absence of an apolipoprotein E-
4 allele), and levels of education. The residual survival times of the subjects are assumed
to follow a parametric Weibull distribution that depends on two covariates age at entry
and the APOE4 status. We further assume that a shared random intercept connects these
two components transition and survival within the same subject.

Recall that the model likelihood function is constructed as follows

$$L = \int \prod_{i=1}^{N} \prod_{l=2}^{n_i} \prod_{\substack{s=1\dots3, \\ v=1\dots4}} (P_{sv}(\boldsymbol{\Theta}|\boldsymbol{X},\boldsymbol{\gamma}))^{\delta y_{i,l-1},s \delta y_{i,l},v} \times \pi_{S_i}(\boldsymbol{\Theta}|\boldsymbol{Z},\boldsymbol{\gamma})dF(\boldsymbol{\gamma})$$

Suppose there are N subjects in the study. The random vector $\boldsymbol{Y}_i = (Y_{i1},\dots,Y_{in_i})$
represents the cognitive states for the ith subject at n_i different ordered assessments. We
let P_{sv} denote the one step transition probability from state s to v. Here $P_{sv} = P_{sv}^{(m)}$ is
dependent on time/age m. The residual survival time of the subject is assumed to follow a
Weibull distribution. π_{S_i} denotes the probability of the ith subject with survival time S_i

failing from the competing risk of death. P_{sv} and π_{S_i} are determined by the commonly known risk factors for dementia and death, respectively. The dependence of the two components within the same subject is captured by the shared random effect γ. Integrating over the whole support of the random effect, we thus obtain the likelihood function displayed above.

Again the resultant integral can be approximated numerically using the Gauss quadrature method by assuming a normally distributed random effect with mean 0 and unknown variance σ^2. This part of the calculation is done using the SAS NLMIXED procedure. The maximum likelihood estimates are produced and presented in Table 3.1 below.

Table 3.1 Parameter estimates for transition probabilities and residual survival in the Nun Study (Initial state: Intact cognition)

Risk Factor	MCI	GI	Dementia	Residual Survival time
Prior State:				
Intact Cognition	-1.1598*	-3.7610*	-5.2779*	
MCI	0.7218*	-2.3254*	-1.9468*	
AGE	0.0949*	0.1746*	0.1723*	
EDUC < 16 yrs	1.5443*	1.5919*	1.4343*	
EDUC = 16 yrs	0.4723*	0.5041*	0.3792	
APOE4 STATUS	0.8422*	1.2971*	1.4184*	0.4238*
AGE at ENTRY				-0.2017*

*Significant at $P < 0.05$

Estimate of rate = 5.24; Estimate of sigma = 0.94

Table 3.2 Number of dementia and death by age in the Nun Study

APOE4 −, EDUC ≥ 16 yrs, BASELINE = 1

(Total = 117)

BEFORE AGE	80	85	90	95	95+
n_1 (# of DEM)	0	2	3	5	2
n_2 (# of DEAD)	0	12	29	22	5
$n_1/(n_1 + n_2)$	0	0.14	0.09	0.19	0.29
$\frac{n_1}{n} / \frac{n_2}{n}$	0	0.17	0.10	0.23	0.40

APOE4 −, EDUC ≥ 16 yrs, BASELINE = 1, 2, 3

(Total = 330)

BEFORE AGE	80	85	90	95	95+
n_1 (# of DEM)	3	15	26	38	20
n_2 (# of DEAD)	3	25	61	55	23
$n_1/(n_1 + n_2)$	0.5	0.375	0.30	0.41	0.47
$\frac{n_1}{n} / \frac{n_2}{n}$	1	0.60	0.43	0.69	0.87

APOE4 −, EDUC ≥ 16 yrs, BASELINE = 1, 2, 3, 4

(Total = 361)

BEFORE AGE	80	85	90	95	95+
n_1 (# of DEM)	7	26	36	42	22
n_2 (# of DEAD)	3	25	61	55	23
$n_1/(n_1 + n_2)$	0.7	0.51	0.37	0.43	0.49
$\frac{n_1}{n} / \frac{n_2}{n}$	2.33	1.04	0.59	0.76	0.96

Table 3.2 above summarizes the number of dementia and death before age 75, 80, 85, 90, 95 and after among the largest subgroup (the low risk group: negative APOE4, college or above Education) in the Nun Study data. For the baseline intact nuns (baseline=1), the number of cases of dementia observed in the age intervals are small (no larger than 5). Including the baseline M.C.I. and G.I. (baseline=2,3), the value of $\frac{n_1}{n_1+n_2}$ decreases from age 85 to 90 and increases after 90. To further investigate the age effect, we add in a quadratic form of age to the original covariate vector and re-fit the model. It shows that the non-linear effect from age on transition is marginal since the associated Z-normal scores are only marginally significant at 5% level. These tables give a general idea how the probability of dementia before death and relative risk change with age in the data.

Now consider the all subjects case. The empirical distribution of the probability of dementia before death in our data shows: in the low risk group, where the subjects have both risk factors being absent (negative APOE4 w/ high Education), the probability of dementia before death increases with age; while in the high risk groups, where the subjects have at least one of the risk factors being present (positive APOE4 w/ high Education, or negative APOE4 w/ low Education, or positive APOE4 w/ low Education), no significant age effect was found. The results are summarized as in Table 3.3 below.

Table 3.3 Summary of the age effect in the Nun Study (All subjects)

	Pr of Dem before Death	Educ	
		≤ 12 yrs	≥ 16 yrs
Apoe4	+	Insufficient subjects	No age effect
	−	No age effect	↑ by age

Table 3.4 displays the estimated probabilities of dementia before death by different genetics and education group at age points 75, 80, 85, 90, and 95 with the initial state being intact cognition. These probabilities are calculated by replacing the unknown parameter vector with the corresponding MLEs from table 3.1.

Table 3.4 Probability of dementia before death in the Nun Study

(Initial state: Intact cognition)

	AGE				
	95	90	85	80	75
APOE4 -, EDUC > 16 yrs	0.59	0.53	0.45	0.36	0.28
APOE4 -, EDUC = 16 yrs	0.59	0.54	0.49	0.42	0.35
APOE4 -, EDUC < 16 yrs	0.61	0.58	0.56	0.54	0.50
APOE4 +, EDUC > 16 yrs	0.57	0.56	0.57	0.58	0.55
APOE4 +, EDUC = 16 yrs	0.56	0.56	0.58	0.60	0.60
APOE4 +, EDUC < 16 yrs	0.56	0.57	0.61	0.66	0.69

The result indicates that for intact cognitive subjects, having an apolipoprotein E-4 present and low education increases the chance of dementia before death. In particular,

(1) Subjects with APOE4 status being negative:

The probability of converting to dementia before death increases with age. That is, a person is more likely to develop dementia at an elderly age. Education could help protect subjects from getting dementia. A person with lower education level would have greater chance of developing dementia than a same-aged person with higher education level. The education effect diminishes with the increase of age. More specifically, at age 75 the probability of dementia before death associated with less than 16 years of education is 0.50, almost twice of that associated with over 16 years of education (0.28); while at age 95 the difference is only 0.02.

(2) Subjects with APOE4 status being positive:

In general, subjects with positive APOE4 are much more likely to develop dementia at early ages (75, 80, and 85). A person at age 80 with positive APOE4 is almost same or more likely to convert to dementia as a person at age 95 with negative APOE4. The probability of dementia before death decreases with age for subjects having 16 years or less education. Among the highest education group, the change of the probabilities fluctuates by time and no monotone trend shows. The highly educated also retain lower risks of getting dementia except for those at age 95, when the probability is 0.57 slightly larger compared with the lower education levels (0.56). However, the by-age difference is smaller in this case.

The estimated probabilities and odds ratios of dementia before death are computed by replacing the parameter vector with the corresponding maximum likelihood estimates from our model. For odds ratios, the comparison is made on the risk factor APOE (positive versus negative) for the same age and education group. The 90% confidence intervals for the probabilities and odds ratios are estimated using the Delta method as we discussed in the previous section. The results are presented in Figure 3.4 and 3.5.

Figure 3.4 displays the estimated probabilities of developing dementia before death by APOE4, education, and at different age (75, 80, 85, 90, 95) with the initial state being intact cognition. The results show that in general the probabilities are only slightly changed by the genetics or education level in the elderly age groups (age 90 and age 95). In contrast the probabilities are more likely to be affected by these risk factors for nuns starting at a younger age. For nuns at age of 75-85, having an APOE4 present and low education increases the risk of getting dementia before death, and such influence from the genetics or education declines consistently with age. Particularly, in the youngest age group (age 75), the probability of developing dementia before death with APOE4 positive and lower than or equal to 12 years of education is more than twice higher than that with APOE4 negative and higher than 16 years of education.

Figure 3.5 gives the estimated odds ratios of dementia before death in the two contrast APOE4 groups by education level and at age points (75, 80, 85, 90, 95) with the initial state being intact cognition. The results illustrate that having high education could somehow increase the odds ratio on APOE4, especially in the younger age groups.

Figure 3.4 Probabilities and 90% confidence intervals of dementia before death

Initial state: intact cognition; age 75 (blue), age 80 (red), age 85 (cyan), age 90 (brown), and age 95 (purple)

APOE4−, Ed>16

APOE4−, Ed=16

APOE4−, Ed<=12

APOE4+, Ed>16

APOE4+, Ed=16
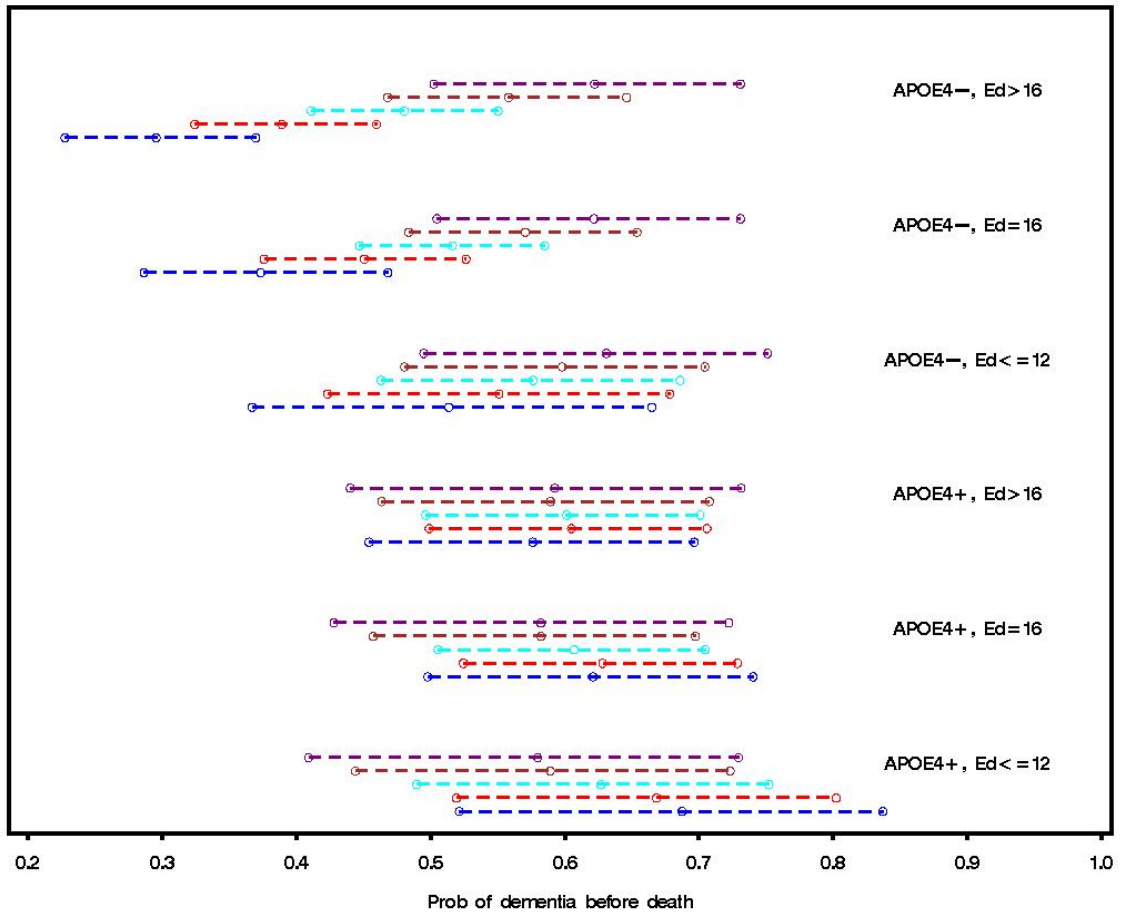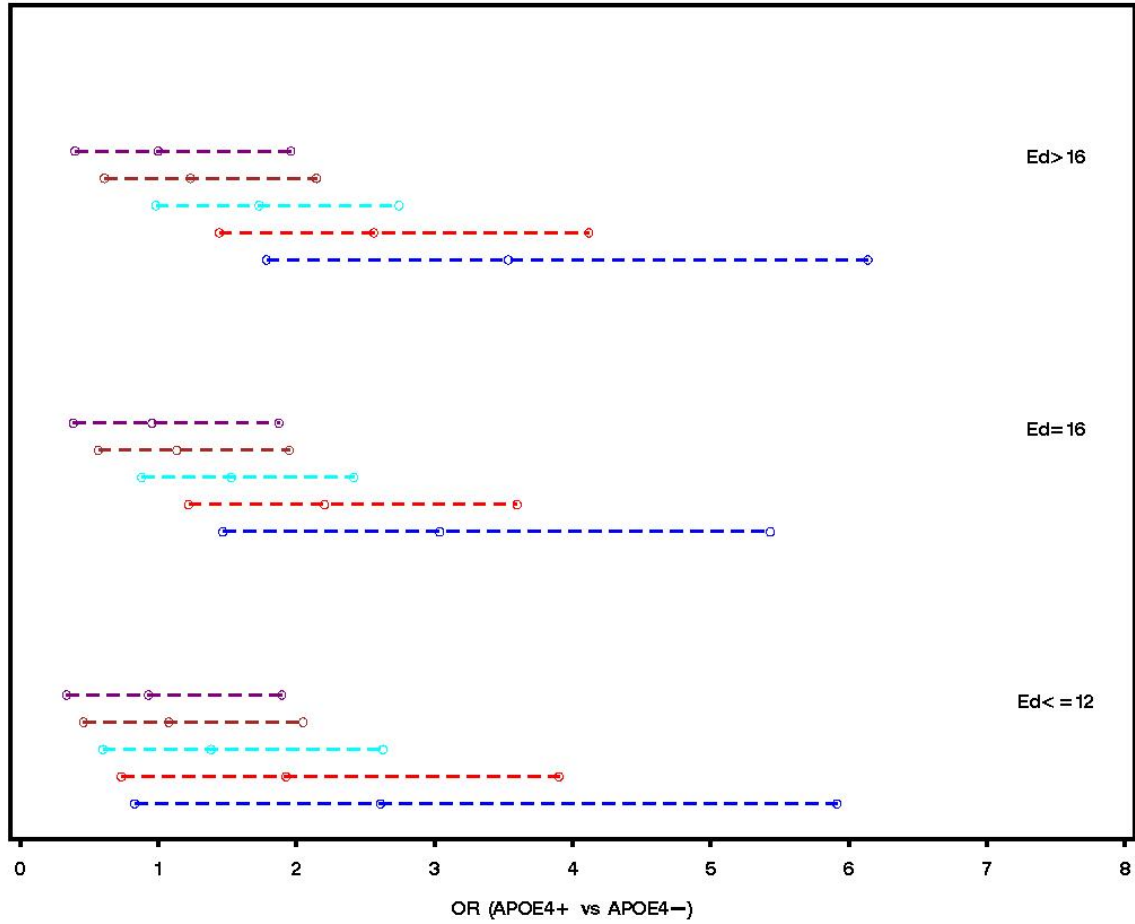
APOE4+, Ed<=12

Prob of dementia before death

Figure 3.5 Odds ratios and 90% confidence intervals of dementia before death (Risk
factor: APOE4+ versus APOE4-)

Initial state: intact cognition; age 75 (blue), age 80 (red), age 85 (cyan), age 90 (brown), and age 95 (purple)

3.5 Risk model

The idea of replacing the original risk factors by a new indicator that represents the presence or absence status of certain risk factors for a given subject came directly from the results in the last section. As we found that among the subgroup of negative APOE4 plus college or above education, which contains the majority of the participants in the study (n=330), there is a clear age effect on the probability of transition to dementia over time in contrast to the other three subgroups. Also, we hope it would help eliminate the observation insufficiency for low education by combining those subgroups with the risk indicator (n=4 for positive APOE4 w/ low education; n=44 for negative APOE4 w/ low education; and n=83 for positive APOE4 w/ high education). It is therefore of interest to compare the results under this "reduced" risk model and to look at the interval estimates of probabilities of developing dementia before death, which would now give more focus on the AGE effect.

Base on the results from Chapter 2, we know that having positive APOE4 and low education are associated with higher risk of transition to dementia before death. We then consider grouping together the original risk factors APOE4 and EDUC by: first, re-define EDUC to be binary variable valued at 0 if having College or above education and 1 otherwise; and second, newly define variable RISK as an indicator of the present/absent risk status of APOE4 and EDUC.

Several models were compared which seemed substantively reasonable:

Model I.   As comparison, the main effects model with risk factors APOE4 (0 = negative; 1 = positive) and EDUC (0 = College or above; 1 = otherwise)

Model II.   The main effects model with a two-level RISK (0 = have neither of the risk factors present; 1 = have at least one of the two risk factors present)

Model III.   The main effects model with a three-level RISK (0 = have neither of the risk factors present; 1 = have either one of the risk factors present; 2 = have both risk factors present) and

Model IV.   The model with 2-way interactions among AGE and RISK (0 = have neither of the risk factors present plus ≤ 85 years of age; 1 = otherwise)

The fit criterions for these models are shown in Table 3.5 as follows. According to AIC/AICC, the three main effects models fit better than the interactions model. The BIC suggests that either the main effects model with a two-level RISK or the main effects model with a three-level RISK are reasonable; given this we opted for the simpler model − Model II, for ease of interpretation and parsimony. The model was fitted to the Nun study data and the results are presented in Table 3.6. The application found that APOE4 is not a significant predictor for survival time. So it was excluded from the final RISK model.

Table 3.5 Comparison of risk models on different fit criterions

Risk Model	-2LogL	AIC	AICC	BIC
Main effects model (apoe4, educ)	5692.5	5736.5	5736.9	5827.4
Main effects model (2-level risk)	5699.0	5737.0	5737.3	5815.5
Main effects model (3-level risk)	5698.3	5736.3	5736.7	5814.9
Interactions model (age*risk)	5722.3	5760.3	5760.7	5838.9

Table 3.6 Maximum likelihood estimates (SE) of model parameters in the Nun Study

(Risk model; Base state: 1 = Intact Cognition)

Risk Factors	State	Estimates	Std. Error	Risk Factors	State	Estimates	Std. Error
Markov chain				Weibull survival			
Age	2	0.099*	(0.016)	Age at Entry	-	-0.204*	(0.021)
	3	0.177*	(0.020)				
	4	0.172*	(0.023)	Rate	-	5.203*	(0.306)
Risk	2	0.899*	(0.199)				
	3	1.222*	(0.228)				
	4	1.299*	(0.260)				
Prior states:				Sigma	-	0.941*	(0.106)
Intact Cognition	2	-1.236*	(0.333)				
	3	-3.830*	(0.325)				
	4	-5.300*	(0.544)				
Mild Cognitive	2	0.691*	(0.326)				
Impairment	3	-2.354*	(0.305)				
	4	-1.963*	(0.326)				

* Significant at $P < 0.05$

We applied the Delta method again to produce the interval estimates for the probabilities of transition to dementia under the risk model. The simulated samples were generated to examine for normality and the results were quite similar to what we obtained under the previous model. The covariance matrix associated with the transition probabilities can be formulated analogously. The estimated mean probabilities with the 90% confidence intervals are displayed in Figure 3.6. These results are consistent with what we found in Section 3.3. Basically age plays an essential role on the transition probability to dementia for those subjects who have no potential genetics or education risk of the disease. In contrast, the subjects who have one or both of such risks show high probability of disease

at an earlier age of 75 while the impact of ageing is slight within the group. The probability of developing dementia increases steadily with age in the low risk group, and it will reach the same level of disease probability (p=0.6) at age 95, which is in average 20 years later than those in the high risk group.

Figure 3.6 Probabilities and 90% confidence intervals of dementia before death (Risk model)

Initial state: intact cognition; age 75 (blue), age 80 (red), age 85 (cyan), age 90 (brown), and age 95 (purple)
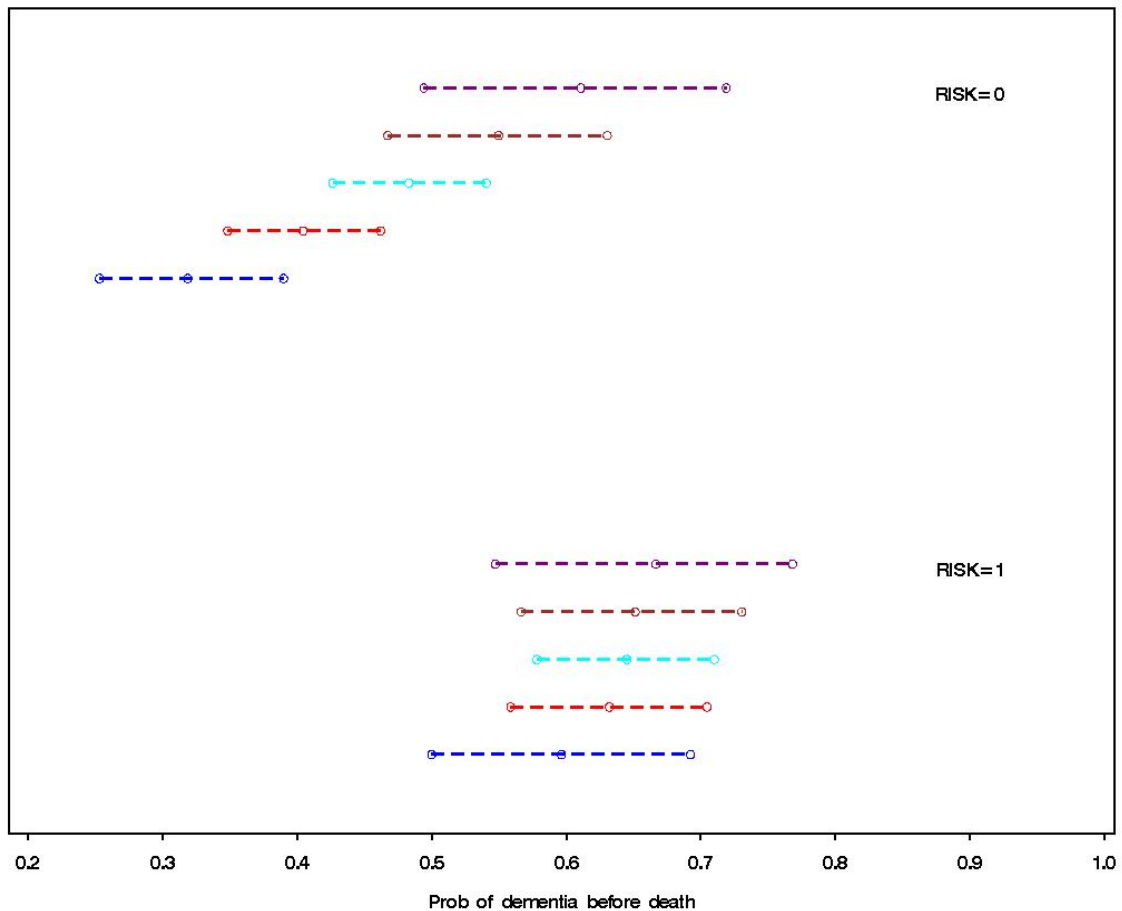
3.6 Conclusion

In the research of chronic diseases it is often of interest to study the absorption statistics that characterize the progression of disease associated with the effects of risk factors. Previous work showed that such statistics always exist in a time homogeneous Markov chain and they are directly related to the fundamental matrix. While in the nonhomogeneous case, proper assumptions are required to ensure the convergence of the corresponding matrix series. This chapter is focused on addressing the problem by accommodating the residual survival time of the subjects confounding in a nonhomogeneous Markov chain. The convergence status of the chain is examined and confirmed. Provide this condition, formulas are derived in computing (1) the probability of developing dementia before death and (2) the 90% confidence interval based on the Delta method.

The results are illustrated with an application to the Nun study. The maximum likelihood estimates of the parameters (presented in Chapter 2) are applied to the computation of the absorption statistics and their confidence intervals. The analysis results indicate that age, education level, and APOE4 status are significant predictors of the transition probabilities. Low education and the presence of APOE4 increase the risk of converting to dementia before death. But the effect diminishes when the age increases. In addition, the odds ratio of dementia before death on APOE4 status (positive versus negative) increases as the level of education increases among younger age groups (age 75, 80, 85).

Chapter 4 : Summary and discussions

4.1 Discussion of the model

In medical applications, the response may refer to a disease state, and this multi-state process is often assumed to be Markov, which greatly simplifies the computation of the likelihood. We developed an improved parametric multi-state Markov model to model this type of longitudinal categorical response data.

The model is based on Salazar's multi-state Markov model (2007). Rather than defining death as a competing absorbing state to dementia, we treated the survival from death as a continuous variable. More specifically, we assumed that the residual survival time S_i of the ith subject follows a Weibull distribution $S_i \sim Weibull(r > 0, \mu_i = e^{\eta_0 + \boldsymbol{\eta}' Z_i + \gamma_i})$, and the probability of this subject failing from the risk of death has the form:

$$\pi(S_i = t_i | \boldsymbol{\Theta}, \boldsymbol{\gamma}) = \left[r e^{\eta_0 + \boldsymbol{\eta}' Z_i + \gamma_i} t_i^{r-1} exp\left(-e^{\eta_0 + \boldsymbol{\eta}' Z_i + \gamma_i} t_i^r\right) \right]^{\tau_i} \left[exp\left(-e^{\eta_0 + \boldsymbol{\eta}' Z_i + \gamma_i} t_i^r\right) \right]^{1-\tau_i} \quad (4.1)$$

Here γ_i is a shared random effect, τ_i is some indicator function valued at 1 if the subject died at time t_i and 0 otherwise, and $\boldsymbol{\Theta}$ represents the vector of all the unknown parameters.

The joint distribution of the categorical response vector $\boldsymbol{Y}_i$ conditional on the baseline state was determined by the product of the conditional distributions of $\boldsymbol{Y}_{ij}$ given $\boldsymbol{Y}_{ij-1}$, assuming the first order Markov property holds. A multinomial logit parameterization could be applied to link these transition probabilities with the fixed and random risk factors, which are expressed as follows:

$$Logit\ P(Y_{ij} = v \mid Y_{ij-1}) = \alpha_v + \boldsymbol{X}_{ij}\boldsymbol{\beta}_v + c_v Y_{ij-1} + \gamma_i \quad for\ v = 2,3,4 \tag{4.2}$$

Base on equations (4.1) and (4.2), the contribution of the ith participant to the likelihood function is

$$L_i(\boldsymbol{\Theta}|\boldsymbol{X}_i, \boldsymbol{Z}_i) = \int \prod_{l=2}^{n_i} \times \prod_{\substack{s=1\cdots3,\\v=1\cdots4}} \left(P_{sv}(\boldsymbol{\Theta}|\boldsymbol{X}_i, \boldsymbol{\gamma}_i)\right)^{\delta_{y_{il-1},s}\delta_{y_{il},v}} \times \pi_{S_i}(\boldsymbol{\Theta}|\boldsymbol{Z}_i, \boldsymbol{\gamma}_i)dF(\boldsymbol{\gamma}_i) \tag{4.3}$$

The functions π and P denote the left hand sides of (4.1) and (4.2) while F denotes the cumulative distribution function for the shared random effect. The overall likelihood can be derived by evaluating the product of (4.3) across the N cohort participants under study. The resultant function was approximated using the Gauss-Hermite quadrature method to produce the maximum likelihood estimates of the parameters.

The evaluation of the likelihood function of the proposed model requires a good choice for the distributional form of the survival term. Simulations in Chapter 2 showed that the model estimates were not sensitive to violations of the Weibull assumption in the case when lognormal is the true prior distribution. We considered four different case scenarios, under which two sets of comparisons were made to investigate: first, the effects of varying the sample size; and second, the effects of misspecification of the distributional form of survival from a possible alternative. The results indicated that when the prior distribution was correctly specified, increasing sample size would help improve the estimates in terms of reducing bias and MSE. On the other hand, only MSE was significantly affected by sample size if the distributional form of survival was misspecified.

4.2 Discussion of the transition probability

To further investigate the long run behavior of the process, we considered the well known canonical form of the one-step transition matrix given by

$$P^{(m,m+1)} = \begin{bmatrix} Q^{(m,m+1)} & R^{(m,m+1)} \\ 0 & I \end{bmatrix}$$

We let m being the initial starting time of the process. Suppose that there are r absorbing states and t transient states in the chain, then $Q^{(m,m+1)}$ is a square $t \times t$ matrix of one-step transition probabilities among the transient states, $R^{(m,m+1)}$ is a $t \times r$ matrix of one-step transition probabilities from a transient state to an absorbing state. For a homogeneous Markov chain, the fundamental matrix $M = (I - Q)^{-1}$ is well-defined and its elements can be calculated explicitly from the converging series

$$M = I + Q + Q^2 + \cdots$$

While in the nonhomogeneous case the fundamental matrix is replaced by the infinite series

$$M^{(m)} = I + Q^{(m,m+1)} + Q^{(m,m+1)}Q^{(m+1,m+2)} + \cdots + Q^{(m,m+1)} \cdots Q^{(m+n-1,m+n)} + \cdots$$

By accommodating the residual life time of the subjects confounding in the chain, we derived the following formula for computing the probability of dementia before death for a given subject assuming initially started the process from state i and baseline time m:

$$P(W < S | \boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_{t^*=1}^{\infty} [Q^{(m,m+t^*-1)} R^{(m+t^*-1,m+t^*)}](i) \cdot [1 - F_S(t^*|\boldsymbol{\theta}, \boldsymbol{\gamma})] \qquad for\ i = 1,2,3$$

Here W is the minimum time it takes for the process to enter an absorbing state, and $F_S(\cdot \mid \boldsymbol{\Theta}, \boldsymbol{\gamma})$ is the cumulative distribution function of survival time S.

Yu et al. (2009) proved that under certain conditions the norm of the substochastic matrix $\left\| \boldsymbol{Q}^{(m,n)} \right\|$ converge to zero as $n \to \infty$, which does not depend on the initial time m or the states. Based on Platis et al.'s sufficient condition of convergence (1998) and the property of cumulative distribution functions, we showed that the probabilities of dementia before death converge with time.

Another primary research interest in the Nun Study is to estimate the confidence intervals associated with the probabilities and odds ratios of developing dementia before death. Our approach by using the Delta method to estimate the corresponding standard errors and to construct the confidence intervals was introduced and illustrated in detail in Chapter 3. Such a technique is based upon the assumption of the asymptotic normality of the sampling distribution of the statistics. The assumption was carefully examined with large simulated samples.

These results for nonhomogeneous Markov chains make it possible to study the effects of the risk factors on the long run behavior of the chain and in the process account for the impact of the competing event death.

4.3 Areas of future research

The model proposed in this dissertation is likelihood based subject specific model conditional on the unobserved latent variables represented by the random effects. One appealing feature of the polychotomous logistic regression model with shared random effect approach is that it allows us to perform statistical inference for the risk factors by means of odds ratios. By fitting the model we were able to study the different roles of the predictors on a subject specific transition.

Further extensions of the model may include that of allowing less strict assumptions about the association between the longitudinal responses and the time-to-event data. The general idea is to define separate but correlated latent variables $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ for $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, and let $f(\boldsymbol{b}_1, \boldsymbol{b}_2)$ denote the joint density, often being bivariate normal. By assuming conditional independence of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ given $(\boldsymbol{b}_1, \boldsymbol{b}_2)$, the joint density of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ can be obtained as

$$f(\boldsymbol{y}_1, \boldsymbol{y}_2) = \iint f(\boldsymbol{y}_1|\boldsymbol{b}_1)f(\boldsymbol{y}_2|\boldsymbol{b}_2)f(\boldsymbol{b}_1, \boldsymbol{b}_2)d\boldsymbol{b}_1 d\boldsymbol{b}_2.$$

The dissertation looked at the impact of violations of the distributional assumption of the survival term over the parameter estimation through simulations. The results are conditional on other aspects of the model specifications such as the mean structure, random effects structure, and linkage function. It is therefore of interest to compare such parametric survival approach with a nonparametric or semiparametric likelihood approach for survival. A more flexible setting is to consider using proportional hazard model for the residual survival time of the subject depending on fixed effects $\boldsymbol{z}_i$ and random effect b_{2i}

$$h_i(t|b_{2i}) = h_0(t)\exp(\mathbf{z}_i'\beta + \alpha' b_{2i})$$

with piecewise-constant baseline hazard (step-functions): $h_0(t) = \sum_{q=1}^{Q} \xi_q I(v_{q-1} < t \le v_q)$ where $0 = v_0 < v_1 < \cdots < v_Q$ denotes a split of the time scale.

In addition, research interests in this area can also be focused on misspecification of the linkage function of the model and procedures to assess goodness-of-fit for multi-state Markov models. Further investigation of the related model stability and verification of the model assumptions are both non-ignorable.

APPENDICES

Appendix A: SAS code for model fitting

```
libname thesis 'C:\Doc\My Doc_research';

ods output ParameterEstimates=est CovMatParmEst=cov;

proc nlmixed data=nun4_M qpoints=1 cov;

parms int1=0.01 int2=0.01 int3=0.01
      age1=0.05 age2=0.05 age3=0.05
      apo1=0.08 apo2=0.08 apo3=0.08
      coll1=0.05 coll2=0.05 coll3=0.05
      grad1=0.05 grad2=0.05 grad3=0.05
      pri11=0.08 pri12=0.08 pri13=0.08
      pri21=0.05 pri22=0.05 pri23=0.05

      intb1=0.01 intb2=0.01 intb3=0.01
      ageb1=0.05 ageb2=0.05 ageb3=0.05
      apob1=0.08 apob2=0.08 apob3=0.08
      collB1=0.05 collB2=0.05 collB3=0.05
      gradB1=0.05 gradB2=0.05 gradB3=0.05

      intc=0.01 entagec=0.05 apoc=0.05
      ratc=0.5
      sd=0.5;

eta1=int1+age1*agec+apo1*apoe4+coll1*ed12+grad1*ed3+pri11*prior1+
pri21*prior2+u;
eta2=int2+age2*agec+apo2*apoe4+coll2*ed12+grad2*ed3+pri12*prior1+
pri22*prior2+u;
eta3=int3+age3*agec+apo3*apoe4+coll3*ed12+grad3*ed3+pri13*prior1+
pri23*prior2+u;

exp_eta1=exp(eta1);
exp_eta2=exp(eta2);
exp_eta3=exp(eta3);
den_eta=1+exp_eta1+ exp_eta2+exp_eta3;

etab1=intb1+ageb1*agec+apob1*apoe4+collB1*ed12+gradB1*ed3+u;
etab2=intb2+ageb2*agec+apob2*apoe4+collB2*ed12+gradB2*ed3+u;
etab3=intb3+ageb3*agec+apob3*apoe4+collB3*ed12+gradB3*ed3+u;

exp_etab1=exp(etab1);
exp_etab2=exp(etab2);
exp_etab3=exp(etab3);
den_etab=1+exp_etab1+ exp_etab2+exp_etab3;
```

```
/* p1 denotes the baseline component */
if (priorstate=1) then p1=(1/den_etab)**baseline;
else if (priorstate=2) then p1=(exp_etab1/den_etab)**baseline;
else if (priorstate=3) then p1=(exp_etab2/den_etab)**baseline;
else if (priorstate=4) then p1=(exp_etab3/den_etab)**baseline;

/* p2 denotes the main transition process */
if (currentstate=1) then p2=(1/den_eta)**ind;
else if (currentstate=2) then p2=(exp_eta1/den_eta)**ind;
else if (currentstate=3) then p2=(exp_eta2/den_eta)**ind;
else if (currentstate=4) then p2=(exp_eta3/den_eta)**ind;

etac=intc+entagec*entrage+apoc*apoe4+u;
exp_etac=exp(etac);

/* p3 denotes the survival component */
if (indxi=1) then p3=(ratc*exp_etac*(survival**(ratc-1))*exp(-
exp_etac*(survival**ratc)))**(baseline*ind);
else if (indxi=0) then p3=(exp(-
exp_etac*(survival**ratc)))**(baseline*ind);

ll=log(p1*p2*p3);

model currentstate ~ general (ll);
random u ~ normal(0,sd*sd) subject=newid;
run;


data thesis.est; set est; run;
data thesis.cov; set cov; run;
```

Appendix B: SAS macro data generation for simulations

```
%MACRO data_gen(m,maxv);

proc iml;

START ptran(AGE, APOE, PRIOR1, PRIOR2, error);
theta={-1.69 0.81 -1.05 0.07 0.19 0.21 0.27 0.39 1.62 0.58 -3.12
-4.15 1.79 -2.43 -2.26};
agec=age-86; /* centered age */
p=j(4,4,0);

p[1,1]=1/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[10]*PR
IOR1+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[11]*PR
IOR1+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[12]*PR
IOR1+error));
p[1,2]=exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[10]*PRIOR1+
error)/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[10]*PRIO
R1+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[11]*PRIO
R1+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[12]*PRIO
R1+error));
p[1,3]=exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[11]*PRIOR1+
error)/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[10]*PRIO
R1+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[11]*PRIO
R1+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[12]*PRIO
R1+error));
p[1,4]=exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[12]*PRIOR1+
error)/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[10]*PRIO
R1+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[11]*PRIO
R1+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[12]*PRIO
R1+error));
p[2,1]=1/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[13]*PR
IOR2+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[14]*PR
IOR2+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[15]*PR
IOR2+error));
p[2,2]=exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[13]*PRIOR2+
error)/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[13]*PRIO
R2+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[14]*PRIO
R2+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[15]*PRIO
R2+error));
p[2,3]=exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[14]*PRIOR2+
error)/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[13]*PRIO
R2+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[14]*PRIO
R2+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[15]*PRIO
R2+error));
p[2,4]=exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[15]*PRIOR2+
error)/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+theta[13]*PRIO
R2+error)+exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+theta[14]*PRIO
R2+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+theta[15]*PRIO
R2+error));
```

```
p[3,1]=1/(1+exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+error)+exp(t
heta[2]+theta[5]*AGEC+theta[8]*APOE+error)+exp(theta[3]+theta[6]*
AGEC+theta[9]*APOE+error));
p[3,2]=exp(theta[1]+theta[4]*AGEC+theta[7]*APOE+error)/(1+exp(the
ta[1]+theta[4]*AGEC+theta[7]*APOE+error)+exp(theta[2]+theta[5]*AG
EC+theta[8]*APOE+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+
error));
p[3,3]=exp(theta[2]+theta[5]*AGEC+theta[8]*APOE+error)/(1+exp(the
ta[1]+theta[4]*AGEC+theta[7]*APOE+error)+exp(theta[2]+theta[5]*AG
EC+theta[8]*APOE+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+
error));
p[3,4]=exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+error)/(1+exp(the
ta[1]+theta[4]*AGEC+theta[7]*APOE+error)+exp(theta[2]+theta[5]*AG
EC+theta[8]*APOE+error)+exp(theta[3]+theta[6]*AGEC+theta[9]*APOE+
error));
p[4,4]=1;

return(p);
FINISH ptran;

start RANDMULTINOMIAL(N, NumTrials, Prob);
    mP = rowvec(Prob);
     d = ncol(mP);
     if N<1 then do;
     print "The requested number of observations should be at
     least 1:" N; stop;
     end;
     if NumTrials <1 then do;
     print "The number of trials should be at least 1:"
     NumTrials; stop;
     end;
     if abs(1 - sum(Prob))>1e-8 then do;
    print "The probabilities must sum to 1:" (sum(Prob))
     [label="Sum"]; stop;
    end;
     if ncol(loc(Prob>0)) < d then do;
     print "Each probability should be positive:" Prob; stop;
     end;
     b = mP;
     order = d + 1 - rank(mP);
     mP[order] = b;
     X = j(n,d,0);
     z = 0;
     do i = 1 to N;
           if d = 1 then do;
                X[i] = NumTrials;
           end;
           else do;
               m = NumTrials;
               q = 1;
               call randgen(z,'BINOM',m,mP[1]);
                 X[i,1] = z;
```

```
                do j = 2 to d-1 by 1 while ( m > 0 );
                m = m - X[i,j-1];
                        q = q - mP[j-1];
                        newp = mP[j]/q;
                        call randgen(z,'BINOM',m,newp);
                        X[i,j] = z;
                end;
                X[i,d] = m - z ;
            end;
        end;
        outX = X;
        outX[ , 1:d] = X[, order];
        return(outX);
finish;


history=j(&m,&maxv,.);
currentage=j(&m,&maxv,.);
apoe4=j(&m,&maxv,.);
survival=j(&m,&maxv,.);
do i=1 to &m;
                error=rannor(1234)*1; /* assign random effects */
                apoe=ranbin(0,1,.2);
                storeapoe=apoe;
                rate=5.08;
                theta_surv={-14.27 0.26 -0.72};
                age=80;surv=0;
                do until (surv>age);
                  age=rannor(0)*3+80;
                  if age>81 then entrage=1; else entrage=0;

                  para=exp(theta_surv[1]+theta_surv[2]*apoe+theta
                  _surv[3]*entrage+error);
                  unisurv=ranuni(0);
                  surv=(1/(para##(1/rate)))*((-
                  log(unisurv))##(1/rate))+75;
                end;
                storesurv=surv;
                storeage=age;
                deltage=exp(0.18);
                prior=1;
                storehist=prior;

    do j=1 to (&maxv-1);
                age=age+deltage;
                        if prior=1 then prior1=1; else prior1=0;
                        if prior=2 then prior2=1; else prior2=0;
                        p=ptran(AGE, APOE, PRIOR1, PRIOR2, error);
                        prob=t(p[prior,]);
                        if prob={0,0,0,1} then current=4;
                        else current=RANDMULTINOMIAL(1, 1,
                        prob)*{1,2,3,4};
```

75

```
                            storeage=storeage // age;
                            storeapoe=storeapoe // apoe;
                            storesurv=storesurv // surv;
                            storehist=storehist // current;
                            prior=current;
        end;

    history=history // t(storehist);
    currentage=currentage // t(storeage);
    apoe4=apoe4//t(storeapoe);
    survival=survival // t(storesurv);
    end;
        apoe4=apoe4[(&m+1):nrow(apoe4),];
        survival=survival[(&m+1):nrow(survival),];
        history=history[(&m+1):nrow(history),];
        currentage=round(currentage[(&m+1):nrow(currentage),],0.001);
          id=t(1:&m);
         do h=1 to &m;
            do k=1 to (ncol(history)-1);
                if history[h,k]=5 then do;
                 history[h,k+1:ncol(history)]=.;
                 currentage[h,k+1:ncol(currentage)]=.;
                 apoe4[h,k+1:ncol(apoe4)]=.;
                 survival[h,k+1:ncol(survival)]=.;
                end;
                if (history[h,k]=4 & (history[h,k+1]=1 |
                 history[h,k+1]=2 | history[h,k+1]=3)) then do;
                 history[h,k+1:ncol(history)]=.;
                 currentage[h,k+1:ncol(currentage)]=.;
                 survival[h,k+1:ncol(survival)]=.;
                end;
            end;
         end;

    data=id || history || currentage || apoe4 || survival;

    dataset1=id || history;
    dataset2=id || currentage ;
    dataset3=id || apoe4;
    dataset4=id || survival;

    create tmp1 from dataset1; append from dataset1;
    create tmp2 from dataset2; append from dataset2;
    create tmp3 from dataset3; append from dataset3;
    create tmp4 from dataset4; append from dataset4;

    proc sort data=tmp1;  by COL1;
    proc sort data=tmp2;  by COL1;
    proc sort data=tmp3;  by COL1;
    proc sort data=tmp4;  by COL1;
```

```sas
data a;
retain id state; drop COL1;
set tmp1;
id=COL1;
array xx(i) COL2-COL50;
do i = 1 to 49;
state=xx;
output;
end;run;

data b;
retain id age;drop COL1;
set tmp2;
id=COL1;
array xx(i) COL2-COL50;
do i = 1 to 49;
age=xx;
output;
end;run;

data c;
retain id apoe; drop COL1;
set tmp3;
id=COL1;
array xx(i) COL2-COL50;
do i = 1 to 49;
apoe=xx;
output;
end;run;

data d;
retain id survival; drop COL1;
set tmp4;
id=COL1;
array xx(i) COL2-COL50;
do i = 1 to 49;
survival=xx;
output;
end;run;

data randata;
retain seq id state age apoe survival;
merge a b c d; by id;
keep seq id state age apoe survival;
seq+1;
if state^=.;
run;
proc sort; by seq; run;
quit;
%mend;
```

Appendix C: SAS macro random sample generation for normality check

```
*nsim:number of iterations;
*age:actual age of subjects;
*apoe:1=positive,0=negative;
*educ:1<=12yrs,13<=2<=15yrs,3=16yrs,4>=17yrs;
*basestate:state at baseline(1,2,or3),default=1;

libname thesis 'C:\Doc\My Doc_research';

%macro check_norm(nsim, age, apoe, educ, basestate);
proc iml;
use thesis.est; read all into A;
use thesis.cov; read all into B;
theta0=A[,1];
cov_theta0=B[,2:27];
randbeta=randnormal(&nsim,theta0,cov_theta0);

ed12=0;
if &educ=1 then ed12=1;
if &educ=2 then ed12=1;
ed3=0;
if &educ=3 then ed3=1;
ENTAGE=&AGE-82;

prob_est=.;
do ll=1 to &nsim;
thetatmp=randbeta[ll,]`;
theta=thetatmp[1:3] // thetatmp[16:21] // thetatmp[4:15];
theta2=thetatmp[22:25];
sd=thetatmp[26];

sum1=0;
do ii=1 to 100;
AGEC=&AGE-88;
AGED=&AGE-75;
QMAT=I(3);RMATEMP=0*j(3,1);

do jj=1 to ii;
M=j(3,3,0); E=j(3,3,1);
M[1,1]=theta[1]+theta[4]+theta[10]*AGEC+theta[13]*&APOE+theta[16]
*ed12+theta[19]*ed3;
M[1,2]=theta[2]+theta[5]+theta[11]*AGEC+theta[14]*&APOE+theta[17]
*ed12+theta[20]*ed3;
M[1,3]=theta[3]+theta[6]+theta[12]*AGEC+theta[15]*&APOE+theta[18]
*ed12+theta[21]*ed3;
M[2,1]=theta[1]+theta[7]+theta[10]*AGEC+theta[13]*&APOE+theta[16]
*ed12+theta[19]*ed3;
M[2,2]=theta[2]+theta[8]+theta[11]*AGEC+theta[14]*&APOE+theta[17]
*ed12+theta[20]*ed3;
```

```
M[2,3]=theta[3]+theta[9]+theta[12]*AGEC+theta[15]*&APOE+theta[18]
*ed12+theta[21]*ed3;
M[3,1]=theta[1]+theta[10]*AGEC+theta[13]*&APOE+theta[16]*ed12+the
ta[19]*ed3;
M[3,2]=theta[2]+theta[11]*AGEC+theta[14]*&APOE+theta[17]*ed12+the
ta[20]*ed3;
M[3,3]=theta[3]+theta[12]*AGEC+theta[15]*&APOE+theta[18]*ed12+the
ta[21]*ed3;

E=exp(M);
P=j(4,4,0);
p[1,1]= 1/(1+E[1,+]);
p[1,2]= E[1,1]/(1+E[1,+]);
p[1,3]= E[1,2]/(1+E[1,+]);
p[1,4]= E[1,3]/(1+E[1,+]);
p[2,1]= 1/(1+E[2,+]);
p[2,2]= E[2,1]/(1+E[2,+]);
p[2,3]= E[2,2]/(1+E[2,+]);
p[2,4]= E[2,3]/(1+E[2,+]);
p[3,1]= 1/(1+E[3,+]);
p[3,2]= E[3,1]/(1+E[3,+]);
p[3,3]= E[3,2]/(1+E[3,+]);
p[3,4]= E[3,3]/(1+E[3,+]);
p[4,4]=1;

Q=P[1:3,1:3];
R=P[1:3,4];
RMAT=QMAT*R;
QMAT=QMAT*Q;
RMATEMP=RMATEMP+RMAT;
AGEC=AGEC+1;
END;

QMATEMP=QMAT*j(3,1);
mu=exp(theta2[1]+theta2[2]*ENTAGE+theta2[3]*&APOE);
delta_F1=exp(-mu#((AGED+ii)##theta2[4]))-exp(-
mu#((AGED+ii+1)##theta2[4]));
sum_temp1=RMATEMP[&basestate]*delta_F1;
sum1=sum1+sum_temp1;
END;

prob1=sum1;
prob_est=prob_est//prob1;
END;
create output1 from prob_est; append from prob_est;
quit;

%mend;


%ODDS1(nsim=10000, age=80, apoe=1, educ=2, basestate=1);
```

```
data output1;
set output1;
if col1 ne .;
col2=log(col1);
col3=col1/(1-col1);
col4=log(col3);
run;

data output1;
set output1;
P=col1;
lnP=col2;
RR=col3;
lnRR=col4;
arsin_sqrtP=arsin(sqrt(P));
keep P lnP RR lnRR arsin_sqrtP;
run;


goptions htext=1.5;
proc univariate data=output1; var P;
histogram P / normal(color=yellow w=2 percents=20 40 60 80
midpercents)
cfill=blue midpoints=0 to 1.3 by 0.03 cframe=ligr;
inset n normal(ksdpval) / pos=ne format=6.3;
run;

proc univariate data=output1; var arsin_sqrtP;
histogram arsin_sqrtP / normal(color=yellow w=2 percents=20 40 60
80 midpercents)
cfill=blue midpoints=0.5 to 1.4 by 0.02 cframe=ligr;
inset n normal(ksdpval) / pos=ne format=6.3;
run;


proc univariate data=output1;
qqplot P lnP RR lnRR arsin_sqrtP / normal(mu=est sigma=est
color=red l=2)
square cframe=ligr;
run;
```

Appendix D: SAS macro to compute partial derivatives for $P^{(m)}$

```
*age:actual age of subjects;
*apoe:1=positive,0=negative;
*educ:1<=12yrs,13<=2<=15yrs,3=16yrs,4>=17yrs;
*basestate:state at baseline(1,2,or3),default=1;

libname thesis 'C:\Doc\My Doc_research';

%macro PARTIAL1(age, apoe, educ, basestate);
proc iml;
use thesis1.est; read all into A;
use thesis1.cov; read all into B;
theta0=A[,1];
cov_theta0=B[,2:27];
theta=theta0[1:3] // theta0[16:21] // theta0[4:15];
theta2=theta0[22:25];
sd=theta0[26];

ed12=0;
if &educ=1 then ed12=1;
if &educ=2 then ed12=1;
ed3=0;
if &educ=3 then ed3=1;
ENTAGE=&AGE-82;

Q_0=I(3);
AGEC=&AGE-88;
AGED=&AGE-75;
M=j(3,3,0); E=j(3,3,1);
M[1,1]=theta[1]+theta[4]+theta[10]*AGEC+theta[13]*&APOE+theta[16]
*ed12+theta[19]*ed3;
M[1,2]=theta[2]+theta[5]+theta[11]*AGEC+theta[14]*&APOE+theta[17]
*ed12+theta[20]*ed3;
M[1,3]=theta[3]+theta[6]+theta[12]*AGEC+theta[15]*&APOE+theta[18]
*ed12+theta[21]*ed3;
M[2,1]=theta[1]+theta[7]+theta[10]*AGEC+theta[13]*&APOE+theta[16]
*ed12+theta[19]*ed3;
M[2,2]=theta[2]+theta[8]+theta[11]*AGEC+theta[14]*&APOE+theta[17]
*ed12+theta[20]*ed3;
M[2,3]=theta[3]+theta[9]+theta[12]*AGEC+theta[15]*&APOE+theta[18]
*ed12+theta[21]*ed3;
M[3,1]=theta[1]+theta[10]*AGEC+theta[13]*&APOE+theta[16]*ed12+the
ta[19]*ed3;
M[3,2]=theta[2]+theta[11]*AGEC+theta[14]*&APOE+theta[17]*ed12+the
ta[20]*ed3;
M[3,3]=theta[3]+theta[12]*AGEC+theta[15]*&APOE+theta[18]*ed12+the
ta[21]*ed3;

E=exp(M);
P=j(4,4,0);
```

```
p[1,1]= 1/(1+E[1,+]);
p[1,2]= E[1,1]/(1+E[1,+]);
p[1,3]= E[1,2]/(1+E[1,+]);
p[1,4]= E[1,3]/(1+E[1,+]);
p[2,1]= 1/(1+E[2,+]);
p[2,2]= E[2,1]/(1+E[2,+]);
p[2,3]= E[2,2]/(1+E[2,+]);
p[2,4]= E[2,3]/(1+E[2,+]);
p[3,1]= 1/(1+E[3,+]);
p[3,2]= E[3,1]/(1+E[3,+]);
p[3,3]= E[3,2]/(1+E[3,+]);
p[3,4]= E[3,3]/(1+E[3,+]);
p[4,4]=1;
QMAT=P[1:3,1:3];
RMAT=P[1:3,4];
mu=exp(theta2[1]+theta2[2]*ENTAGE+theta2[3]*&APOE);
F1=exp(-mu#((AGED+1)##theta2[4]));
FMAT=F1#I(3);


A=j(9,24,0);
B=j(3,24,0);
C=j(9,24,0);
A[1,1]=    -E[1,1]/((1+E[1,+])##2);
A[2,1]=    -E[2,1]/((1+E[2,+])##2);
A[3,1]=    -E[3,1]/((1+E[3,+])##2);
A[4,1]=    (1+E[1,+]-E[1,1])*E[1,1]/((1+E[1,+])##2);
A[5,1]=    (1+E[2,+]-E[2,1])*E[2,1]/((1+E[2,+])##2);
A[6,1]=    (1+E[3,+]-E[3,1])*E[3,1]/((1+E[3,+])##2);
A[7,1]=    -E[1,2]*E[1,1]/((1+E[1,+])##2);
A[8,1]=    -E[2,2]*E[2,1]/((1+E[2,+])##2);
A[9,1]=    -E[3,2]*E[3,1]/((1+E[3,+])##2);
A[1,2]=    -E[1,2]/((1+E[1,+])##2);
A[2,2]=    -E[2,2]/((1+E[2,+])##2);
A[3,2]=    -E[3,2]/((1+E[3,+])##2);
A[4,2]=    -E[1,1]*E[1,2]/((1+E[1,+])##2);
A[5,2]=    -E[2,1]*E[2,2]/((1+E[2,+])##2);
A[6,2]=    -E[3,1]*E[3,2]/((1+E[3,+])##2);
A[7,2]=    (1+E[1,+]-E[1,2])*E[1,2]/((1+E[1,+])##2);
A[8,2]=    (1+E[2,+]-E[2,2])*E[2,2]/((1+E[2,+])##2);
A[9,2]=    (1+E[3,+]-E[3,2])*E[3,2]/((1+E[3,+])##2);
A[1,3]=    -E[1,3]/((1+E[1,+])##2);
A[2,3]=    -E[2,3]/((1+E[2,+])##2);
A[3,3]=    -E[3,3]/((1+E[3,+])##2);
A[4,3]=    -E[1,1]*E[1,3]/((1+E[1,+])##2);
A[5,3]=    -E[2,1]*E[2,3]/((1+E[2,+])##2);
A[6,3]=    -E[3,1]*E[3,3]/((1+E[3,+])##2);
A[7,3]=    -E[1,2]*E[1,3]/((1+E[1,+])##2);
A[8,3]=    -E[2,2]*E[2,3]/((1+E[2,+])##2);
A[9,3]=    -E[3,2]*E[3,3]/((1+E[3,+])##2);
A[1,4]=    -E[1,1]/((1+E[1,+])##2);
A[4,4]=    (1+E[1,+]-E[1,1])*E[1,1]/((1+E[1,+])##2);
A[7,4]=    -E[1,2]*E[1,1]/((1+E[1,+])##2);
```

```
A[1,5]=    -E[1,2]/((1+E[1,+])##2);
A[4,5]=    -E[1,1]*E[1,2]/((1+E[1,+])##2);
A[7,5]=    (1+E[1,+]-E[1,2])*E[1,2]/((1+E[1,+])##2);
A[1,6]=    -E[1,3]/((1+E[1,+])##2);
A[4,6]=    -E[1,1]*E[1,3]/((1+E[1,+])##2);
A[7,6]=    -E[1,2]*E[1,3]/((1+E[1,+])##2);
A[2,7]=    -E[2,1]/((1+E[2,+])##2);
A[5,7]=    (1+E[2,+]-E[2,1])*E[2,1]/((1+E[2,+])##2);
A[8,7]=    -E[2,2]*E[2,1]/((1+E[2,+])##2);
A[2,8]=    -E[2,2]/((1+E[2,+])##2);
A[5,8]=    -E[2,1]*E[2,2]/((1+E[2,+])##2);
A[8,8]=    (1+E[2,+]-E[2,2])*E[2,2]/((1+E[2,+])##2);
A[2,9]=    -E[2,3]/((1+E[2,+])##2);
A[5,9]=    -E[2,1]*E[2,3]/((1+E[2,+])##2);
A[8,9]=    -E[2,2]*E[2,3]/((1+E[2,+])##2);
A[1,10]=   -E[1,1]*AGEC/((1+E[1,+])##2);
A[2,10]=   -E[2,1]*AGEC/((1+E[2,+])##2);
A[3,10]=   -E[3,1]*AGEC/((1+E[3,+])##2);
A[4,10]=   (1+E[1,+]-E[1,1])*E[1,1]*AGEC/((1+E[1,+])##2);
A[5,10]=   (1+E[2,+]-E[2,1])*E[2,1]*AGEC/((1+E[2,+])##2);
A[6,10]=   (1+E[3,+]-E[3,1])*E[3,1]*AGEC/((1+E[3,+])##2);
A[7,10]=   -E[1,2]*E[1,1]*AGEC/((1+E[1,+])##2);
A[8,10]=   -E[2,2]*E[2,1]*AGEC/((1+E[2,+])##2);
A[9,10]=   -E[3,2]*E[3,1]*AGEC/((1+E[3,+])##2);
A[1,11]=   -E[1,2]*AGEC/((1+E[1,+])##2);
A[2,11]=   -E[2,2]*AGEC/((1+E[2,+])##2);
A[3,11]=   -E[3,2]*AGEC/((1+E[3,+])##2);
A[4,11]=   -E[1,1]*E[1,2]*AGEC/((1+E[1,+])##2);
A[5,11]=   -E[2,1]*E[2,2]*AGEC/((1+E[2,+])##2);
A[6,11]=   -E[3,1]*E[3,2]*AGEC/((1+E[3,+])##2);
A[7,11]=   (1+E[1,+]-E[1,2])*E[1,2]*AGEC/((1+E[1,+])##2);
A[8,11]=   (1+E[2,+]-E[2,2])*E[2,2]*AGEC/((1+E[2,+])##2);
A[9,11]=   (1+E[3,+]-E[3,2])*E[3,2]*AGEC/((1+E[3,+])##2);
A[1,12]=   -E[1,3]*AGEC/((1+E[1,+])##2);
A[2,12]=   -E[2,3]*AGEC/((1+E[2,+])##2);
A[3,12]=   -E[3,3]*AGEC/((1+E[3,+])##2);
A[4,12]=   -E[1,1]*E[1,3]*AGEC/((1+E[1,+])##2);
A[5,12]=   -E[2,1]*E[2,3]*AGEC/((1+E[2,+])##2);
A[6,12]=   -E[3,1]*E[3,3]*AGEC/((1+E[3,+])##2);
A[7,12]=   -E[1,2]*E[1,3]*AGEC/((1+E[1,+])##2);
A[8,12]=   -E[2,2]*E[2,3]*AGEC/((1+E[2,+])##2);
A[9,12]=   -E[3,2]*E[3,3]*AGEC/((1+E[3,+])##2);
A[1,13]=   -E[1,1]*&APOE/((1+E[1,+])##2);
A[2,13]=   -E[2,1]*&APOE/((1+E[2,+])##2);
A[3,13]=   -E[3,1]*&APOE/((1+E[3,+])##2);
A[4,13]=   (1+E[1,+]-E[1,1])*E[1,1]*&APOE/((1+E[1,+])##2);
A[5,13]=   (1+E[2,+]-E[2,1])*E[2,1]*&APOE/((1+E[2,+])##2);
A[6,13]=   (1+E[3,+]-E[3,1])*E[3,1]*&APOE/((1+E[3,+])##2);
A[7,13]=   -E[1,2]*E[1,1]*&APOE/((1+E[1,+])##2);
A[8,13]=   -E[2,2]*E[2,1]*&APOE/((1+E[2,+])##2);
A[9,13]=   -E[3,2]*E[3,1]*&APOE/((1+E[3,+])##2);
A[1,14]=   -E[1,2]*&APOE/((1+E[1,+])##2);
```

```
A[2,14]=   -E[2,2]*&APOE/((1+E[2,+])##2);
A[3,14]=   -E[3,2]*&APOE/((1+E[3,+])##2);
A[4,14]=   -E[1,1]*E[1,2]*&APOE/((1+E[1,+])##2);
A[5,14]=   -E[2,1]*E[2,2]*&APOE/((1+E[2,+])##2);
A[6,14]=   -E[3,1]*E[3,2]*&APOE/((1+E[3,+])##2);
A[7,14]=   (1+E[1,+]-E[1,2])*E[1,2]*&APOE/((1+E[1,+])##2);
A[8,14]=   (1+E[2,+]-E[2,2])*E[2,2]*&APOE/((1+E[2,+])##2);
A[9,14]=   (1+E[3,+]-E[3,2])*E[3,2]*&APOE/((1+E[3,+])##2);
A[1,15]=   -E[1,3]*&APOE/((1+E[1,+])##2);
A[2,15]=   -E[2,3]*&APOE/((1+E[2,+])##2);
A[3,15]=   -E[3,3]*&APOE/((1+E[3,+])##2);
A[4,15]=   -E[1,1]*E[1,3]*&APOE/((1+E[1,+])##2);
A[5,15]=   -E[2,1]*E[2,3]*&APOE/((1+E[2,+])##2);
A[6,15]=   -E[3,1]*E[3,3]*&APOE/((1+E[3,+])##2);
A[7,15]=   -E[1,2]*E[1,3]*&APOE/((1+E[1,+])##2);
A[8,15]=   -E[2,2]*E[2,3]*&APOE/((1+E[2,+])##2);
A[9,15]=   -E[3,2]*E[3,3]*&APOE/((1+E[3,+])##2);
A[1,16]=   -E[1,1]*ed12/((1+E[1,+])##2);
A[2,16]=   -E[2,1]*ed12/((1+E[2,+])##2);
A[3,16]=   -E[3,1]*ed12/((1+E[3,+])##2);
A[4,16]=   (1+E[1,+]-E[1,1])*E[1,1]*ed12/((1+E[1,+])##2);
A[5,16]=   (1+E[2,+]-E[2,1])*E[2,1]*ed12/((1+E[2,+])##2);
A[6,16]=   (1+E[3,+]-E[3,1])*E[3,1]*ed12/((1+E[3,+])##2);
A[7,16]=   -E[1,2]*E[1,1]*ed12/((1+E[1,+])##2);
A[8,16]=   -E[2,2]*E[2,1]*ed12/((1+E[2,+])##2);
A[9,16]=   -E[3,2]*E[3,1]*ed12/((1+E[3,+])##2);
A[1,17]=   -E[1,2]*ed12/((1+E[1,+])##2);
A[2,17]=   -E[2,2]*ed12/((1+E[2,+])##2);
A[3,17]=   -E[3,2]*ed12/((1+E[3,+])##2);
A[4,17]=   -E[1,1]*E[1,2]*ed12/((1+E[1,+])##2);
A[5,17]=   -E[2,1]*E[2,2]*ed12/((1+E[2,+])##2);
A[6,17]=   -E[3,1]*E[3,2]*ed12/((1+E[3,+])##2);
A[7,17]=   (1+E[1,+]-E[1,2])*E[1,2]*ed12/((1+E[1,+])##2);
A[8,17]=   (1+E[2,+]-E[2,2])*E[2,2]*ed12/((1+E[2,+])##2);
A[9,17]=   (1+E[3,+]-E[3,2])*E[3,2]*ed12/((1+E[3,+])##2);
A[1,18]=   -E[1,3]*ed12/((1+E[1,+])##2);
A[2,18]=   -E[2,3]*ed12/((1+E[2,+])##2);
A[3,18]=   -E[3,3]*ed12/((1+E[3,+])##2);
A[4,18]=   -E[1,1]*E[1,3]*ed12/((1+E[1,+])##2);
A[5,18]=   -E[2,1]*E[2,3]*ed12/((1+E[2,+])##2);
A[6,18]=   -E[3,1]*E[3,3]*ed12/((1+E[3,+])##2);
A[7,18]=   -E[1,2]*E[1,3]*ed12/((1+E[1,+])##2);
A[8,18]=   -E[2,2]*E[2,3]*ed12/((1+E[2,+])##2);
A[9,18]=   -E[3,2]*E[3,3]*ed12/((1+E[3,+])##2);
A[1,19]=   -E[1,1]*ed3/((1+E[1,+])##2);
A[2,19]=   -E[2,1]*ed3/((1+E[2,+])##2);
A[3,19]=   -E[3,1]*ed3/((1+E[3,+])##2);
A[4,19]=   (1+E[1,+]-E[1,1])*E[1,1]*ed3/((1+E[1,+])##2);
A[5,19]=   (1+E[2,+]-E[2,1])*E[2,1]*ed3/((1+E[2,+])##2);
A[6,19]=   (1+E[3,+]-E[3,1])*E[3,1]*ed3/((1+E[3,+])##2);
A[7,19]=   -E[1,2]*E[1,1]*ed3/((1+E[1,+])##2);
A[8,19]=   -E[2,2]*E[2,1]*ed3/((1+E[2,+])##2);
```

```
A[9,19]=   -E[3,2]*E[3,1]*ed3/((1+E[3,+])##2);
A[1,20]=   -E[1,2]*ed3/((1+E[1,+])##2);
A[2,20]=   -E[2,2]*ed3/((1+E[2,+])##2);
A[3,20]=   -E[3,2]*ed3/((1+E[3,+])##2);
A[4,20]=   -E[1,1]*E[1,2]*ed3/((1+E[1,+])##2);
A[5,20]=   -E[2,1]*E[2,2]*ed3/((1+E[2,+])##2);
A[6,20]=   -E[3,1]*E[3,2]*ed3/((1+E[3,+])##2);
A[7,20]=   (1+E[1,+]-E[1,2])*E[1,2]*ed3/((1+E[1,+])##2);
A[8,20]=   (1+E[2,+]-E[2,2])*E[2,2]*ed3/((1+E[2,+])##2);
A[9,20]=   (1+E[3,+]-E[3,2])*E[3,2]*ed3/((1+E[3,+])##2);
A[1,21]=   -E[1,3]*ed3/((1+E[1,+])##2);
A[2,21]=   -E[2,3]*ed3/((1+E[2,+])##2);
A[3,21]=   -E[3,3]*ed3/((1+E[3,+])##2);
A[4,21]=   -E[1,1]*E[1,3]*ed3/((1+E[1,+])##2);
A[5,21]=   -E[2,1]*E[2,3]*ed3/((1+E[2,+])##2);
A[6,21]=   -E[3,1]*E[3,3]*ed3/((1+E[3,+])##2);
A[7,21]=   -E[1,2]*E[1,3]*ed3/((1+E[1,+])##2);
A[8,21]=   -E[2,2]*E[2,3]*ed3/((1+E[2,+])##2);
A[9,21]=   -E[3,2]*E[3,3]*ed3/((1+E[3,+])##2);
B[1,1]=    -E[1,3]*E[1,1]/((1+E[1,+])##2);
B[2,1]=    -E[2,3]*E[2,1]/((1+E[2,+])##2);
B[3,1]=    -E[3,3]*E[3,1]/((1+E[3,+])##2);
B[1,2]=    -E[1,3]*E[1,2]/((1+E[1,+])##2);
B[2,2]=    -E[2,3]*E[2,2]/((1+E[2,+])##2);
B[3,2]=    -E[3,3]*E[3,2]/((1+E[3,+])##2);
B[1,3]=    (1+E[1,+]-E[1,3])*E[1,3]/((1+E[1,+])##2);
B[2,3]=    (1+E[2,+]-E[2,3])*E[2,3]/((1+E[2,+])##2);
B[3,3]=    (1+E[3,+]-E[3,3])*E[3,3]/((1+E[3,+])##2);
B[1,4]=    -E[1,3]*E[1,1]/((1+E[1,+])##2);
B[1,5]=    -E[1,3]*E[1,2]/((1+E[1,+])##2);
B[1,6]=    (1+E[1,+]-E[1,3])*E[1,3]/((1+E[1,+])##2);
B[2,7]=    -E[2,3]*E[2,1]/((1+E[2,+])##2);
B[2,8]=    -E[2,3]*E[2,2]/((1+E[2,+])##2);
B[2,9]=    (1+E[2,+]-E[2,3])*E[2,3]/((1+E[2,+])##2);
B[1,10]=   -E[1,3]*E[1,1]*AGEC/((1+E[1,+])##2);
B[2,10]=   -E[2,3]*E[2,1]*AGEC/((1+E[2,+])##2);
B[3,10]=   -E[3,3]*E[3,1]*AGEC/((1+E[3,+])##2);
B[1,11]=   -E[1,3]*E[1,2]*AGEC/((1+E[1,+])##2);
B[2,11]=   -E[2,3]*E[2,2]*AGEC/((1+E[2,+])##2);
B[3,11]=   -E[3,3]*E[3,2]*AGEC/((1+E[3,+])##2);
B[1,12]=   (1+E[1,+]-E[1,3])*E[1,3]*AGEC/((1+E[1,+])##2);
B[2,12]=   (1+E[2,+]-E[2,3])*E[2,3]*AGEC/((1+E[2,+])##2);
B[3,12]=   (1+E[3,+]-E[3,3])*E[3,3]*AGEC/((1+E[3,+])##2);
B[1,13]=   -E[1,3]*E[1,1]*&APOE/((1+E[1,+])##2);
B[2,13]=   -E[2,3]*E[2,1]*&APOE/((1+E[2,+])##2);
B[3,13]=   -E[3,3]*E[3,1]*&APOE/((1+E[3,+])##2);
B[1,14]=   -E[1,3]*E[1,2]*&APOE/((1+E[1,+])##2);
B[2,14]=   -E[2,3]*E[2,2]*&APOE/((1+E[2,+])##2);
B[3,14]=   -E[3,3]*E[3,2]*&APOE/((1+E[3,+])##2);
B[1,15]=   (1+E[1,+]-E[1,3])*E[1,3]*&APOE/((1+E[1,+])##2);
B[2,15]=   (1+E[2,+]-E[2,3])*E[2,3]*&APOE/((1+E[2,+])##2);
B[3,15]=   (1+E[3,+]-E[3,3])*E[3,3]*&APOE/((1+E[3,+])##2);
```

```
B[1,16]=   -E[1,3]*E[1,1]*ed12/((1+E[1,+])##2);
B[2,16]=   -E[2,3]*E[2,1]*ed12/((1+E[2,+])##2);
B[3,16]=   -E[3,3]*E[3,1]*ed12/((1+E[3,+])##2);
B[1,17]=   -E[1,3]*E[1,2]*ed12/((1+E[1,+])##2);
B[2,17]=   -E[2,3]*E[2,2]*ed12/((1+E[2,+])##2);
B[3,17]=   -E[3,3]*E[3,2]*ed12/((1+E[3,+])##2);
B[1,18]=   (1+E[1,+]-E[1,3])*E[1,3]*ed12/((1+E[1,+])##2);
B[2,18]=   (1+E[2,+]-E[2,3])*E[2,3]*ed12/((1+E[2,+])##2);
B[3,18]=   (1+E[3,+]-E[3,3])*E[3,3]*ed12/((1+E[3,+])##2);
B[1,19]=   -E[1,3]*E[1,1]*ed3/((1+E[1,+])##2);
B[2,19]=   -E[2,3]*E[2,1]*ed3/((1+E[2,+])##2);
B[3,19]=   -E[3,3]*E[3,1]*ed3/((1+E[3,+])##2);
B[1,20]=   -E[1,3]*E[1,2]*ed3/((1+E[1,+])##2);
B[2,20]=   -E[2,3]*E[2,2]*ed3/((1+E[2,+])##2);
B[3,20]=   -E[3,3]*E[3,2]*ed3/((1+E[3,+])##2);
B[1,21]=   (1+E[1,+]-E[1,3])*E[1,3]*ed3/((1+E[1,+])##2);
B[2,21]=   (1+E[2,+]-E[2,3])*E[2,3]*ed3/((1+E[2,+])##2);
B[3,21]=   (1+E[3,+]-E[3,3])*E[3,3]*ed3/((1+E[3,+])##2);
C[1,22]= (-mu#((AGED+1)##theta2[4]))*exp(mu#((AGED+1)
##theta2[4]));
C[5,22]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]));
C[9,22]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]));
C[1,23]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*ENTAGE;
C[5,23]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*ENTAGE;
C[9,23]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*ENTAGE;
C[1,24]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*&APOE;
C[5,24]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*&APOE;
C[9,24]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*&APOE;
AMAT=A;
BMAT=B;
CMAT=C;
partemp=(RMAT`@I(3))*CMAT+FMAT*BMAT;
AGEC=AGEC+1;

do k=2 to 100;
M=j(3,3,0); E=j(3,3,1);
M[1,1]=theta[1]+theta[4]+theta[10]*AGEC+theta[13]*&APOE+theta[16]
*ed12+theta[19]*ed3;
M[1,2]=theta[2]+theta[5]+theta[11]*AGEC+theta[14]*&APOE+theta[17]
*ed12+theta[20]*ed3;
M[1,3]=theta[3]+theta[6]+theta[12]*AGEC+theta[15]*&APOE+theta[18]
*ed12+theta[21]*ed3;
M[2,1]=theta[1]+theta[7]+theta[10]*AGEC+theta[13]*&APOE+theta[16]
*ed12+theta[19]*ed3;
```

```
M[2,2]=theta[2]+theta[8]+theta[11]*AGEC+theta[14]*&APOE+theta[17]
*ed12+theta[20]*ed3;
M[2,3]=theta[3]+theta[9]+theta[12]*AGEC+theta[15]*&APOE+theta[18]
*ed12+theta[21]*ed3;
M[3,1]=theta[1]+theta[10]*AGEC+theta[13]*&APOE+theta[16]*ed12+the
ta[19]*ed3;
M[3,2]=theta[2]+theta[11]*AGEC+theta[14]*&APOE+theta[17]*ed12+the
ta[20]*ed3;
M[3,3]=theta[3]+theta[12]*AGEC+theta[15]*&APOE+theta[18]*ed12+the
ta[21]*ed3;
E=exp(M);
P=j(4,4,0);
p[1,1]= 1/(1+E[1,+]);
p[1,2]= E[1,1]/(1+E[1,+]);
p[1,3]= E[1,2]/(1+E[1,+]);
p[1,4]= E[1,3]/(1+E[1,+]);
p[2,1]= 1/(1+E[2,+]);
p[2,2]= E[2,1]/(1+E[2,+]);
p[2,3]= E[2,2]/(1+E[2,+]);
p[2,4]= E[2,3]/(1+E[2,+]);
p[3,1]= 1/(1+E[3,+]);
p[3,2]= E[3,1]/(1+E[3,+]);
p[3,3]= E[3,2]/(1+E[3,+]);
p[3,4]= E[3,3]/(1+E[3,+]);
p[4,4]=1;
Q=P[1:3,1:3];
R=P[1:3,4];
QMAT=QMAT // Q;
RMAT=RMAT // R;
mu=exp(theta2[1]+theta2[2]*ENTAGE+theta2[3]*&APOE);
F1=exp(-mu#((AGED+k)##theta2[4]));
F=F1#I(3);
FMAT=FMAT // F;

A=j(9,24,0);
B=j(3,24,0);
C=j(9,24,0);
A[1,1]=    -E[1,1]/((1+E[1,+])##2);
A[2,1]=    -E[2,1]/((1+E[2,+])##2);
A[3,1]=    -E[3,1]/((1+E[3,+])##2);
A[4,1]=    (1+E[1,+]-E[1,1])*E[1,1]/((1+E[1,+])##2);
A[5,1]=    (1+E[2,+]-E[2,1])*E[2,1]/((1+E[2,+])##2);
A[6,1]=    (1+E[3,+]-E[3,1])*E[3,1]/((1+E[3,+])##2);
A[7,1]=    -E[1,2]*E[1,1]/((1+E[1,+])##2);
A[8,1]=    -E[2,2]*E[2,1]/((1+E[2,+])##2);
A[9,1]=    -E[3,2]*E[3,1]/((1+E[3,+])##2);
A[1,2]=    -E[1,2]/((1+E[1,+])##2);
A[2,2]=    -E[2,2]/((1+E[2,+])##2);
A[3,2]=    -E[3,2]/((1+E[3,+])##2);
A[4,2]=    -E[1,1]*E[1,2]/((1+E[1,+])##2);
A[5,2]=    -E[2,1]*E[2,2]/((1+E[2,+])##2);
A[6,2]=    -E[3,1]*E[3,2]/((1+E[3,+])##2);
```

```
A[7,2]=    (1+E[1,+]-E[1,2])*E[1,2]/((1+E[1,+])##2);
A[8,2]=    (1+E[2,+]-E[2,2])*E[2,2]/((1+E[2,+])##2);
A[9,2]=    (1+E[3,+]-E[3,2])*E[3,2]/((1+E[3,+])##2);
A[1,3]=    -E[1,3]/((1+E[1,+])##2);
A[2,3]=    -E[2,3]/((1+E[2,+])##2);
A[3,3]=    -E[3,3]/((1+E[3,+])##2);
A[4,3]=    -E[1,1]*E[1,3]/((1+E[1,+])##2);
A[5,3]=    -E[2,1]*E[2,3]/((1+E[2,+])##2);
A[6,3]=    -E[3,1]*E[3,3]/((1+E[3,+])##2);
A[7,3]=    -E[1,2]*E[1,3]/((1+E[1,+])##2);
A[8,3]=    -E[2,2]*E[2,3]/((1+E[2,+])##2);
A[9,3]=    -E[3,2]*E[3,3]/((1+E[3,+])##2);
A[1,4]=    -E[1,1]/((1+E[1,+])##2);
A[4,4]=    (1+E[1,+]-E[1,1])*E[1,1]/((1+E[1,+])##2);
A[7,4]=    -E[1,2]*E[1,1]/((1+E[1,+])##2);
A[1,5]=    -E[1,2]/((1+E[1,+])##2);
A[4,5]=    -E[1,1]*E[1,2]/((1+E[1,+])##2);
A[7,5]=    (1+E[1,+]-E[1,2])*E[1,2]/((1+E[1,+])##2);
A[1,6]=    -E[1,3]/((1+E[1,+])##2);
A[4,6]=    -E[1,1]*E[1,3]/((1+E[1,+])##2);
A[7,6]=    -E[1,2]*E[1,3]/((1+E[1,+])##2);
A[2,7]=    -E[2,1]/((1+E[2,+])##2);
A[5,7]=    (1+E[2,+]-E[2,1])*E[2,1]/((1+E[2,+])##2);
A[8,7]=    -E[2,2]*E[2,1]/((1+E[2,+])##2);
A[2,8]=    -E[2,2]/((1+E[2,+])##2);
A[5,8]=    -E[2,1]*E[2,2]/((1+E[2,+])##2);
A[8,8]=    (1+E[2,+]-E[2,2])*E[2,2]/((1+E[2,+])##2);
A[2,9]=    -E[2,3]/((1+E[2,+])##2);
A[5,9]=    -E[2,1]*E[2,3]/((1+E[2,+])##2);
A[8,9]=    -E[2,2]*E[2,3]/((1+E[2,+])##2);
A[1,10]=   -E[1,1]*AGEC/((1+E[1,+])##2);
A[2,10]=   -E[2,1]*AGEC/((1+E[2,+])##2);
A[3,10]=   -E[3,1]*AGEC/((1+E[3,+])##2);
A[4,10]=   (1+E[1,+]-E[1,1])*E[1,1]*AGEC/((1+E[1,+])##2);
A[5,10]=   (1+E[2,+]-E[2,1])*E[2,1]*AGEC/((1+E[2,+])##2);
A[6,10]=   (1+E[3,+]-E[3,1])*E[3,1]*AGEC/((1+E[3,+])##2);
A[7,10]=   -E[1,2]*E[1,1]*AGEC/((1+E[1,+])##2);
A[8,10]=   -E[2,2]*E[2,1]*AGEC/((1+E[2,+])##2);
A[9,10]=   -E[3,2]*E[3,1]*AGEC/((1+E[3,+])##2);
A[1,11]=   -E[1,2]*AGEC/((1+E[1,+])##2);
A[2,11]=   -E[2,2]*AGEC/((1+E[2,+])##2);
A[3,11]=   -E[3,2]*AGEC/((1+E[3,+])##2);
A[4,11]=   -E[1,1]*E[1,2]*AGEC/((1+E[1,+])##2);
A[5,11]=   -E[2,1]*E[2,2]*AGEC/((1+E[2,+])##2);
A[6,11]=   -E[3,1]*E[3,2]*AGEC/((1+E[3,+])##2);
A[7,11]=   (1+E[1,+]-E[1,2])*E[1,2]*AGEC/((1+E[1,+])##2);
A[8,11]=   (1+E[2,+]-E[2,2])*E[2,2]*AGEC/((1+E[2,+])##2);
A[9,11]=   (1+E[3,+]-E[3,2])*E[3,2]*AGEC/((1+E[3,+])##2);
A[1,12]=   -E[1,3]*AGEC/((1+E[1,+])##2);
A[2,12]=   -E[2,3]*AGEC/((1+E[2,+])##2);
A[3,12]=   -E[3,3]*AGEC/((1+E[3,+])##2);
A[4,12]=   -E[1,1]*E[1,3]*AGEC/((1+E[1,+])##2);
```

```
A[5,12]=   -E[2,1]*E[2,3]*AGEC/((1+E[2,+])##2);
A[6,12]=   -E[3,1]*E[3,3]*AGEC/((1+E[3,+])##2);
A[7,12]=   -E[1,2]*E[1,3]*AGEC/((1+E[1,+])##2);
A[8,12]=   -E[2,2]*E[2,3]*AGEC/((1+E[2,+])##2);
A[9,12]=   -E[3,2]*E[3,3]*AGEC/((1+E[3,+])##2);
A[1,13]=   -E[1,1]*&APOE/((1+E[1,+])##2);
A[2,13]=   -E[2,1]*&APOE/((1+E[2,+])##2);
A[3,13]=   -E[3,1]*&APOE/((1+E[3,+])##2);
A[4,13]=   (1+E[1,+]-E[1,1])*E[1,1]*&APOE/((1+E[1,+])##2);
A[5,13]=   (1+E[2,+]-E[2,1])*E[2,1]*&APOE/((1+E[2,+])##2);
A[6,13]=   (1+E[3,+]-E[3,1])*E[3,1]*&APOE/((1+E[3,+])##2);
A[7,13]=   -E[1,2]*E[1,1]*&APOE/((1+E[1,+])##2);
A[8,13]=   -E[2,2]*E[2,1]*&APOE/((1+E[2,+])##2);
A[9,13]=   -E[3,2]*E[3,1]*&APOE/((1+E[3,+])##2);
A[1,14]=   -E[1,2]*&APOE/((1+E[1,+])##2);
A[2,14]=   -E[2,2]*&APOE/((1+E[2,+])##2);
A[3,14]=   -E[3,2]*&APOE/((1+E[3,+])##2);
A[4,14]=   -E[1,1]*E[1,2]*&APOE/((1+E[1,+])##2);
A[5,14]=   -E[2,1]*E[2,2]*&APOE/((1+E[2,+])##2);
A[6,14]=   -E[3,1]*E[3,2]*&APOE/((1+E[3,+])##2);
A[7,14]=   (1+E[1,+]-E[1,2])*E[1,2]*&APOE/((1+E[1,+])##2);
A[8,14]=   (1+E[2,+]-E[2,2])*E[2,2]*&APOE/((1+E[2,+])##2);
A[9,14]=   (1+E[3,+]-E[3,2])*E[3,2]*&APOE/((1+E[3,+])##2);
A[1,15]=   -E[1,3]*&APOE/((1+E[1,+])##2);
A[2,15]=   -E[2,3]*&APOE/((1+E[2,+])##2);
A[3,15]=   -E[3,3]*&APOE/((1+E[3,+])##2);
A[4,15]=   -E[1,1]*E[1,3]*&APOE/((1+E[1,+])##2);
A[5,15]=   -E[2,1]*E[2,3]*&APOE/((1+E[2,+])##2);
A[6,15]=   -E[3,1]*E[3,3]*&APOE/((1+E[3,+])##2);
A[7,15]=   -E[1,2]*E[1,3]*&APOE/((1+E[1,+])##2);
A[8,15]=   -E[2,2]*E[2,3]*&APOE/((1+E[2,+])##2);
A[9,15]=   -E[3,2]*E[3,3]*&APOE/((1+E[3,+])##2);
A[1,16]=   -E[1,1]*ed12/((1+E[1,+])##2);
A[2,16]=   -E[2,1]*ed12/((1+E[2,+])##2);
A[3,16]=   -E[3,1]*ed12/((1+E[3,+])##2);
A[4,16]=   (1+E[1,+]-E[1,1])*E[1,1]*ed12/((1+E[1,+])##2);
A[5,16]=   (1+E[2,+]-E[2,1])*E[2,1]*ed12/((1+E[2,+])##2);
A[6,16]=   (1+E[3,+]-E[3,1])*E[3,1]*ed12/((1+E[3,+])##2);
A[7,16]=   -E[1,2]*E[1,1]*ed12/((1+E[1,+])##2);
A[8,16]=   -E[2,2]*E[2,1]*ed12/((1+E[2,+])##2);
A[9,16]=   -E[3,2]*E[3,1]*ed12/((1+E[3,+])##2);
A[1,17]=   -E[1,2]*ed12/((1+E[1,+])##2);
A[2,17]=   -E[2,2]*ed12/((1+E[2,+])##2);
A[3,17]=   -E[3,2]*ed12/((1+E[3,+])##2);
A[4,17]=   -E[1,1]*E[1,2]*ed12/((1+E[1,+])##2);
A[5,17]=   -E[2,1]*E[2,2]*ed12/((1+E[2,+])##2);
A[6,17]=   -E[3,1]*E[3,2]*ed12/((1+E[3,+])##2);
A[7,17]=   (1+E[1,+]-E[1,2])*E[1,2]*ed12/((1+E[1,+])##2);
A[8,17]=   (1+E[2,+]-E[2,2])*E[2,2]*ed12/((1+E[2,+])##2);
A[9,17]=   (1+E[3,+]-E[3,2])*E[3,2]*ed12/((1+E[3,+])##2);
A[1,18]=   -E[1,3]*ed12/((1+E[1,+])##2);
A[2,18]=   -E[2,3]*ed12/((1+E[2,+])##2);
```

```
A[3,18]=    -E[3,3]*ed12/((1+E[3,+])##2);
A[4,18]=    -E[1,1]*E[1,3]*ed12/((1+E[1,+])##2);
A[5,18]=    -E[2,1]*E[2,3]*ed12/((1+E[2,+])##2);
A[6,18]=    -E[3,1]*E[3,3]*ed12/((1+E[3,+])##2);
A[7,18]=    -E[1,2]*E[1,3]*ed12/((1+E[1,+])##2);
A[8,18]=    -E[2,2]*E[2,3]*ed12/((1+E[2,+])##2);
A[9,18]=    -E[3,2]*E[3,3]*ed12/((1+E[3,+])##2);
A[1,19]=    -E[1,1]*ed3/((1+E[1,+])##2);
A[2,19]=    -E[2,1]*ed3/((1+E[2,+])##2);
A[3,19]=    -E[3,1]*ed3/((1+E[3,+])##2);
A[4,19]=    (1+E[1,+]-E[1,1])*E[1,1]*ed3/((1+E[1,+])##2);
A[5,19]=    (1+E[2,+]-E[2,1])*E[2,1]*ed3/((1+E[2,+])##2);
A[6,19]=    (1+E[3,+]-E[3,1])*E[3,1]*ed3/((1+E[3,+])##2);
A[7,19]=    -E[1,2]*E[1,1]*ed3/((1+E[1,+])##2);
A[8,19]=    -E[2,2]*E[2,1]*ed3/((1+E[2,+])##2);
A[9,19]=    -E[3,2]*E[3,1]*ed3/((1+E[3,+])##2);
A[1,20]=    -E[1,2]*ed3/((1+E[1,+])##2);
A[2,20]=    -E[2,2]*ed3/((1+E[2,+])##2);
A[3,20]=    -E[3,2]*ed3/((1+E[3,+])##2);
A[4,20]=    -E[1,1]*E[1,2]*ed3/((1+E[1,+])##2);
A[5,20]=    -E[2,1]*E[2,2]*ed3/((1+E[2,+])##2);
A[6,20]=    -E[3,1]*E[3,2]*ed3/((1+E[3,+])##2);
A[7,20]=    (1+E[1,+]-E[1,2])*E[1,2]*ed3/((1+E[1,+])##2);
A[8,20]=    (1+E[2,+]-E[2,2])*E[2,2]*ed3/((1+E[2,+])##2);
A[9,20]=    (1+E[3,+]-E[3,2])*E[3,2]*ed3/((1+E[3,+])##2);
A[1,21]=    -E[1,3]*ed3/((1+E[1,+])##2);
A[2,21]=    -E[2,3]*ed3/((1+E[2,+])##2);
A[3,21]=    -E[3,3]*ed3/((1+E[3,+])##2);
A[4,21]=    -E[1,1]*E[1,3]*ed3/((1+E[1,+])##2);
A[5,21]=    -E[2,1]*E[2,3]*ed3/((1+E[2,+])##2);
A[6,21]=    -E[3,1]*E[3,3]*ed3/((1+E[3,+])##2);
A[7,21]=    -E[1,2]*E[1,3]*ed3/((1+E[1,+])##2);
A[8,21]=    -E[2,2]*E[2,3]*ed3/((1+E[2,+])##2);
A[9,21]=    -E[3,2]*E[3,3]*ed3/((1+E[3,+])##2);
B[1,1]=     -E[1,3]*E[1,1]/((1+E[1,+])##2);
B[2,1]=     -E[2,3]*E[2,1]/((1+E[2,+])##2);
B[3,1]=     -E[3,3]*E[3,1]/((1+E[3,+])##2);
B[1,2]=     -E[1,3]*E[1,2]/((1+E[1,+])##2);
B[2,2]=     -E[2,3]*E[2,2]/((1+E[2,+])##2);
B[3,2]=     -E[3,3]*E[3,2]/((1+E[3,+])##2);
B[1,3]=     (1+E[1,+]-E[1,3])*E[1,3]/((1+E[1,+])##2);
B[2,3]=     (1+E[2,+]-E[2,3])*E[2,3]/((1+E[2,+])##2);
B[3,3]=     (1+E[3,+]-E[3,3])*E[3,3]/((1+E[3,+])##2);
B[1,4]=     -E[1,3]*E[1,1]/((1+E[1,+])##2);
B[1,5]=     -E[1,3]*E[1,2]/((1+E[1,+])##2);
B[1,6]=     (1+E[1,+]-E[1,3])*E[1,3]/((1+E[1,+])##2);
B[2,7]=     -E[2,3]*E[2,1]/((1+E[2,+])##2);
B[2,8]=     -E[2,3]*E[2,2]/((1+E[2,+])##2);
B[2,9]=     (1+E[2,+]-E[2,3])*E[2,3]/((1+E[2,+])##2);
B[1,10]=    -E[1,3]*E[1,1]*AGEC/((1+E[1,+])##2);
B[2,10]=    -E[2,3]*E[2,1]*AGEC/((1+E[2,+])##2);
B[3,10]=    -E[3,3]*E[3,1]*AGEC/((1+E[3,+])##2);
```

```
B[1,11]=   -E[1,3]*E[1,2]*AGEC/((1+E[1,+])##2);
B[2,11]=   -E[2,3]*E[2,2]*AGEC/((1+E[2,+])##2);
B[3,11]=   -E[3,3]*E[3,2]*AGEC/((1+E[3,+])##2);
B[1,12]=   (1+E[1,+]-E[1,3])*E[1,3]*AGEC/((1+E[1,+])##2);
B[2,12]=   (1+E[2,+]-E[2,3])*E[2,3]*AGEC/((1+E[2,+])##2);
B[3,12]=   (1+E[3,+]-E[3,3])*E[3,3]*AGEC/((1+E[3,+])##2);
B[1,13]=   -E[1,3]*E[1,1]*&APOE/((1+E[1,+])##2);
B[2,13]=   -E[2,3]*E[2,1]*&APOE/((1+E[2,+])##2);
B[3,13]=   -E[3,3]*E[3,1]*&APOE/((1+E[3,+])##2);
B[1,14]=   -E[1,3]*E[1,2]*&APOE/((1+E[1,+])##2);
B[2,14]=   -E[2,3]*E[2,2]*&APOE/((1+E[2,+])##2);
B[3,14]=   -E[3,3]*E[3,2]*&APOE/((1+E[3,+])##2);
B[1,15]=   (1+E[1,+]-E[1,3])*E[1,3]*&APOE/((1+E[1,+])##2);
B[2,15]=   (1+E[2,+]-E[2,3])*E[2,3]*&APOE/((1+E[2,+])##2);
B[3,15]=   (1+E[3,+]-E[3,3])*E[3,3]*&APOE/((1+E[3,+])##2);
B[1,16]=   -E[1,3]*E[1,1]*ed12/((1+E[1,+])##2);
B[2,16]=   -E[2,3]*E[2,1]*ed12/((1+E[2,+])##2);
B[3,16]=   -E[3,3]*E[3,1]*ed12/((1+E[3,+])##2);
B[1,17]=   -E[1,3]*E[1,2]*ed12/((1+E[1,+])##2);
B[2,17]=   -E[2,3]*E[2,2]*ed12/((1+E[2,+])##2);
B[3,17]=   -E[3,3]*E[3,2]*ed12/((1+E[3,+])##2);
B[1,18]=   (1+E[1,+]-E[1,3])*E[1,3]*ed12/((1+E[1,+])##2);
B[2,18]=   (1+E[2,+]-E[2,3])*E[2,3]*ed12/((1+E[2,+])##2);
B[3,18]=   (1+E[3,+]-E[3,3])*E[3,3]*ed12/((1+E[3,+])##2);
B[1,19]=   -E[1,3]*E[1,1]*ed3/((1+E[1,+])##2);
B[2,19]=   -E[2,3]*E[2,1]*ed3/((1+E[2,+])##2);
B[3,19]=   -E[3,3]*E[3,1]*ed3/((1+E[3,+])##2);
B[1,20]=   -E[1,3]*E[1,2]*ed3/((1+E[1,+])##2);
B[2,20]=   -E[2,3]*E[2,2]*ed3/((1+E[2,+])##2);
B[3,20]=   -E[3,3]*E[3,2]*ed3/((1+E[3,+])##2);
B[1,21]=   (1+E[1,+]-E[1,3])*E[1,3]*ed3/((1+E[1,+])##2);
B[2,21]=   (1+E[2,+]-E[2,3])*E[2,3]*ed3/((1+E[2,+])##2);
B[3,21]=   (1+E[3,+]-E[3,3])*E[3,3]*ed3/((1+E[3,+])##2);
C[1,22]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]));
C[5,22]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]));
C[9,22]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]));
C[1,23]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*ENTAGE;
C[5,23]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*ENTAGE;
C[9,23]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*ENTAGE;
C[1,24]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*&APOE;
C[5,24]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*&APOE;
C[9,24]= (-mu#((AGED+1)##theta2[4]))*exp(-mu#((AGED+1)
##theta2[4]))*&APOE;
```

```
AMAT=AMAT // A;
BMAT=BMAT // B;
CMAT=CMAT // C;
AGEC=AGEC+1;
END;


TEMP_TOTAL=j(3,24,0);
do kk=2 to 100;
    TEMP_A=I(9);
    TEMP_C=I(3);
    TEMP_B=j(9,24,0);
    do ii=1 to kk-1;
        TEMP_B1=I(9);
        TEMP_B2=I(3);
        if ii=kk-1 then TEMP_B1=I(9);
        else do pp=ii+1 to kk-1;
        T_B1=QMAT[(3*pp-2):(3*pp),]`@I(3);
        TEMP_B1=TEMP_B1*T_B1;
        END;
        do qq=1 to ii-1;
        T_B2=QMAT[(3*qq-2):(3*qq),];
        TEMP_B2=TEMP_B2*T_B2;
        END;
    T_A=QMAT[(3*ii-2):(3*ii),]`@I(3);
    TEMP_A=TEMP_A*T_A;
    T_C=QMAT[(3*ii-2):(3*ii),];
    TEMP_C=TEMP_C*T_C;
    S_B=TEMP_B1*(I(3)@(FMAT[(3*kk-2):(3*kk),]*TEMP_B2))
*AMAT[(9*ii-8):(9*ii),];
    TEMP_B=TEMP_B+S_B;
    END;
    TEMP_TS=(RMAT[(3*kk-2):(3*kk),]`@I(3))*(TEMP_A*CMAT[(9*kk-
8):(9*kk),]+TEMP_B)+FMAT[(3*kk-2):(3*kk),]*TEMP_C*BMAT[(3*kk-
2):(3*kk),];
    TEMP_TOTAL=TEMP_TOTAL+TEMP_TS;
END;
PAR_TOTAL=partemp+TEMP_TOTAL;
print PAR_TOTAL;
quit;
%mend;
```

BIBLIOGRAPHY

[1] Agresti, A., *Categorical data analysis, 2nd ed.* John Wiley & Sons Inc., Hoboken. 2002.

[2] Bhat, U. N., Miller, G.K. *Elements of Applied Stochastic Processes.* 3rd ed. Wiley, New York. 2002.

[3] Breslow, N. E., Lin, X., Bias correction in generalized linear mixed models with a single component of dispersion. *Biometrika* 82, 81-91, 1995.

[4] Craig B. A., Sendi P. P., Estimation of the transition matrix of a discrete-time Markov chain. *Health Economics* 11, 33-42, 2002.

[5] Efron, B. Nonparametric estimates of standard error, the jacknife, the bootstrap, and other methods. *Biometrika*, 68:589-599, 1981.

[6] Elashoff R.M., Li G, Li N. An approach to joint analysis of longitudinal measurements and competing risks failure time data. *Stat. Med.* 26:2813-2835, 2007.

[7] Fieuws S. and Verbeke G., Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* 62, 424-431, 2006.

[8] Fitzmaurice, G.M., Laird N.M., A likelihood-based method for analyzing longitudinal binary responses. *Biometrika* 80, 141-151, 1993.

[9] Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., *Longitudinal Data Analysis.* Chapman & Hall/CRC, 2008.

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

[11] Gao, S., A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data. *Statistics in Medicine* 23, 211-219, 2004.

[12] Gibbons, R. D., Hedeker, D., Application of random-effects probit regression models. *Journal of Consulting and Clinical Psychology* 62, 285-296, 1994.

[13] Heagerty, P. J., Marginally specified logistic-normal models for longitudinal binary data. *Biometrics* 55, 688-698, 1999.

[14] Heagerty, P. J., Zeger, S. L., Marginalized multilevel models and likelihood inference. *Statistical Science* 15, 1, 1-19, 2000.

[15] Heagerty, P.J., Kurland, B. F., Misspecified maximum likelihood estimates and generalized linear mixed models. *Biometrika* 88, 973-985, 2001.

[16] Heagerty, P. J., Marginalized transitional models and likelihood inference for longitudinal categorical data. *Biometrics* 58, 342-351, 2002.

[17] Henderson, R., Diggle, P., Dobson A., Joint modelling of longitudinal measurements and event time data. *Biostatistics* 465-480, 2000.

[18] Hogan, J. W. and Laird, N. M., Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine* 16, 239-257, 1997a.

[19] Hogan, J. W. and Laird, N. M., Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine* 16, 259-272, 1997b.

[20] Huang X., Stefanski L.A., Davidian M., Latent-Model Robustness in Joint Models for a Primary Endpoint and a Longitudinal Process. *Biometrics* 65, 719-727, 2009.

[21] Kay, R. A., Markov model for analyzing cancer markers and diseases states in survival studies *Biometrics* 42, 855-865, 1986.

[22] Kryscio R.J., Yu L., Snowdon D., Tyas S.L. Models for identifying risk factors for transitions from intact cognition to mild cognitive impairment and dementia. *Proceedings of the Section on Epidemiology*, Joint Statistical Meetings, Denver, August, 2008.

[23] Lee, K., Daniels M. J., A class of Markov models for longitudinal ordinal data. *Biometrics* 63, 1060-1067, 2007.

[24] Li, J., Yang X., Wu, Y., Shoptaw, S., A random-effects Markov transition model for Poisson-distributed repeated measures with non-ignorable missing values, *Statistics in Medicine*, 26, 2519-2532, 2007.

[25] Li, Y.P., Chan W., Analysis of longitudinal multinomial outcome data, *Biometrical Journal* 48(2), 319-326, 2006.

[26] Liu, Q., Pierce, D. A., A note on Gauss-Hermite quadrature. *Biometrika* 81, 624-629, 1994.

[27] McArdle J.J., Small B.J., Backman L., Fratiglioni L., Longitudinal models of growth and survival applied to the early detection of Alzheimer's disease. *J Geriatric Psychiatry Neurol*, 18:234-241, 2005.

[28] McCullagh, P., Nelder, J.A., *Generalized linear model 2nd ed.* Chapman & Hall, London. 1983.

[29] Muenz L.R., Rubinstein L.V., Markov models for covariate dependence of binary sequences. *Biometrics* 41, 91-101, 1985.

[30] Platis A, Limnios N, Du M.L., Hitting time in a finite non-homogeneous Markov chain with applications, *Applied Stochastic Models and Data Analysis* 14, 241-253, 1998.

[31] Pulkstenis, E. P., Ten Have, T. R., Landis, J.R., Model for the analysis of binary longitudinal data subject to informative drop-out through remedication. *J. Am. Stat. Assoc.* 93, 438-450, 1998.

[32] Rizopoulos, D., Verbeke, G., Molenberghs, G., Shared parameter models under random effects misspecification. *Biometrika* 95, 63-74, 2008.

[33] Rubin, D. B., Inference and missing data. *Biometrika* 63, 581-590, 1976.

[34] Salazar, J.C., Schmitt, F.A., Yu, L., Mendiondo, M.S. Kryscio, R.J. Shared random effects analysis of multi-state Markov models: application to a longitudinal study of transitions to dementia. *Statistics in Medicine,* 26:568-580, 2007.

[35] SAS Institute Inc., *SAS/STAT User's Guide, Version 8*. SAS Institute Inc., Cary, NC, 1999.

[36] Skrondal, A., Rabe-Hesketh, S. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton: Chapman & Hall/CRC, 2004.

[37] Smith, M. K., SAS –proc nlmixed, *Pharmaceutical Statistics* 2, 69-75, 2003.

[38] Snowdon D.A., Greiner L.H., Mortimer J.A., Riley K.P., Greiner P.A., Markesbery W.R., Brain Infarction and the Clinical Expression of Alzheimer Disease: The Nun Study. *JAMA* 277, 813-817, 1997.

[39] Stiratelli, R., Laird, N. M., Ware, J. H., Random-effects models for serial observations with binary response. *Biometrics* 40, 961-971, 1984.

[40] Ten Have, T. R., Kunselman, A., Pulkstenis, E. P., Landis, J. R., Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* 54, 367-383, 1998.

[41] Ten Have, T. R., Miller, M. E., Reboussin, B. A., James, M. K., Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the Longitudinal Study of Aging. *Biometrics* 56, 279-287, 2000.

[42] Ten Have, T. R., Reboussin, B. A., Miller, M. E., Kunselman, A., Mixed effects logistic regression models for multiple longitudinal binary functional limitation response with informative drop-out and confounding by baseline outcomes. *Biometrics* 58, 137-144, 2002.

[43] Tsiatis A.A. and Davidian M., Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 14, 809-834, 2004.

[44] Tyas, S.L., Salazar, J.C., Snowdon, D., Derosiers, M. F., Riley, K.P., Mendiondo, M.S., Kryscio, R. J. Transitions to mild cognitive impairments, dementia, and death: finding from the Nun Study. *Am. J. Epidemiol.* 165:1231-1238, 2007.

[45] Verbeke, G., Lesaffre, E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* 53, 541-556, 1997.

[46] Vonesh, E.F., Greene, T., Schluchter, M.D., Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine* 25, 143-163, 2006.

[47] Wang, Z., Louis, T.A., Marginalized binary mixed effects models with covariatedependent random effects and likelihood inference. *Biometrics* 60, 884-891, 2004.

[48] Wedderburn, R. W. M. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* 61, 439-447, 1974.

[49] Xu, J. and Zeger, S. L. Joint analysis of longitudinal data comprising repeated measures and times to events. *Appl. Statist.* 50:375-387, 2001.

[50] Ye W., Lin X., Taylor J.M., Semiparametric modeling of longitudinal measurements and time-to-event data – a two-stage regression calibration approach. *Biometrics* 64, 1238-1246, 2008.

[51] Yu L., Tyas S.L., Snowdon D., Kryscio R.J. Effects of ignoring baseline on modeling transitions from intact cognition to dementia. *Computational Statistics and Data Analysis.* 53:3334-3343, 2009.

[52] Zeger, S. L. and Karim, M. R., Generalized linear models with random effects: a Gibbs sampling approach. *J. Am. Stat. Assoc.* 86, 79-86, 1991.

Vita

Liou Xu

Date of Birth: 10 February 1981

Place of Birth: Shanghai, China

Education:

2004 - 2006   MS in Statistics, University of Kentucky, Lexington, KY

1999 - 2003   BS in Statistics, Fudan University, Shanghai, CN

Positions and Employment:

06/2006 - Present   Research Assistant, The Biostatistics and Data Management Core, Alzheimer's Disease Research Center (ADRC), University of Kentucky

06/2009 - 09/2009   Intern Biostatistician, Department of Biostatistics and Clinical Science, Sanofi Pasteur

08/2004 - 05/2006   Teaching Assistant, Department of Statistics, College of Arts and Sciences, University of Kentucky

06/2003 – 06/2004   Business Analyst, Advanced Technique Group (ATG), J.Walter Thompson, Shanghai, CN

Publications:

Liou Xu, David Snowdon, Richard J. Kryscio, "A Markov transition model to dementia with death as a competing event", *Proceedings of the Section on Epidemiology*, Joint Statistical Meetings in Washington, DC , August, 2009.

Peter T. Nelson, Richard J. Kryscio, Gregory A. Jicha, Erin L. Abner, Frederick A. Schmitt, Liou Xu, Greg Cooper, Charles D. Smith, William R. Markesbery, "Relative preservation of MMSE scores in autopsy-proven dementia with Lewy bodies", *Neurology*, 2009; 73(14):1127-1133.