Bayesian Joint Modeling of Longitudinal and Time-to-Event Data with Application to Alzheimer's Disease

BY

YAMIN WANG

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Public Health Sciences
in the Graduate College of the
University of Illinois at Chicago, 2021

Chicago, Illinois

Dissertation Committee:

Sanjib Basu, Chair and Advisor

Saria Awadalla, Epidemiology and Biostatistics

Michael Berbaum, Biostatistics

Dulal Bhaumik, Psychiatry, Biostatistics and Bioengineering

Li Liu, Epidemiology and Biostatistics

ACKNOWLEDGEMENTS

I would like to take this opportunity now to express my sincere thanks to the following people who have been supporting and accompanying me over the past years.

First I would like to express my warmest and unreserved thanks to my advisor, Dr. Sanjib Basu, for his consistent guidance and support. I have benefited greatly from Dr. Basu's inspiration, patience and vast knowledge in Bayesian modeling for my doctoral study. I am especially grateful for his sharp comments and suggestions on this research.

I would like to thank other members of my dissertation committee, Dr. Saria Awadalla, Dr. Michael Berbaum, Dr. Dulal Bhaumik, and Dr. Li Liu, for their help during my thesis work. Without their valuable advice and insights, this thesis would not have become possible. I want to especially thank Dr. Berbaum for sharing ideas and providing the references for analyzing changepoint data.

I want to appreciate the all faculty and staff members of the Epidemiology and Biostatistics Division at School of Public Health for their kindness and help that made my graduate studies a wonderful experience. I also want to thank all the fellow students in the division for their companionship along the way.

Most importantly, I am grateful to my family, especially my husband, Weihan, for his endless love, patience, and continuous encourage on me. And I also want to thank my two beautiful daughters, who bring so many joys to me.

My last gratitude must go out to Rush Alzheimer's Disease Center for providing the data for my research. MAP resource can be requested at https://www.radc.rush.edu and MAP project is funded by National Institute on Aging (NIA) grant: R01AG017917.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

CHAPTER PAGE

TABLE OF CONTENTS (CONTINUED)

LIST OF TABLES

LIST OF TABLES (CONTINUED)

LIST OF FIGURES

LIST OF FIGURES (CONTINUED)

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
AFT	Accelerated failure time
ANOVA	Analysis of variance
APOE	Apolipoprotein E
BIC	Bayesian information criterion
BMI	Body mass index
BUGS	Bayesian Inference Using Gibbs Sampling
CDC	Centers for diseases control and prevention
LCA	Latent class analysis
LCM	Latent class modeling
LMEM	Linear mixed effects model
MA	Marquardt algorithm
MAP	Memory and aging project
MCMC	Markov chain Monte Carlo
MLE	Maximum likelihood estimates
MoCA	Montreal cognitive assessment
MMSE	Mini-mental state exam
NIA	National Institute on Aging
PAQUID	Personne Agee Quid
PH	Proportional hazard
RADC	Rush Alzheimer's Disease Center
SD	Standard deviation
SREM	Shared random-effects model

SUMMARY

In many longitudinal epidemiological and clinical studies, it is routine to collect multiple repeated measures, as well as one or multiple time-to-event outcomes. The follow-ups are usually long enough to measure some aspects of the disease. It would be interesting and appropriate to link the longitudinal marker trajectories in a joint model approach in association with the time-to-event outcomes for valid inferences. The joint modeling method has attracted increasing attention in the statistical field recently and many extensions have been explored. In this dissertation, several research topics related to joint modeling of the longitudinal and time-to-event data were investigated. The first part of the work was to construct the Bayesian joint models based on two different linking structures: latent class framework and shared random effects framework. Both single and bivariate longitudinal outcomes were considered. The second part of this work was to propose a joint model with a random changepoint for the non-linear longitudinal marker trajectories. We investigated 5 different formulations to characterize the transition zone for the changepoint data. We further extended the model for a bivariate longitudinal data with correlated changepoints and took into account the competing risks and interval censor in the survival model, which is methodologically challenging. We adopted Bayesian approach for statistical inference and the proposed methodologies were evaluated based on simulation studies. The motivational application for this study is based on the Memory and Aging Project of Rush University Medical Center. Alzheimer's disease, like many other chronic diseases, is a neurodegenerative disease involving a long-term process of cognitive decline and motor dysfunction which often begin before the disease diagnosis. The statistical methodological development in this dissertation aims for a better understanding of the natural history of pre-dementia cognitive aging, motor function change and time to Alzheimer's disease by joint modelling these outcomes together for more insightful and valid statistical inferences.

1 INTRODUCTION

1.1 Joint Statistical Analysis of Longitudinal and Survival Model

In many health-related longitudinal studies, in addition to the primary endpoints for clinical events, repeated measurements of response variables or biomarkers are also collected. Both types of outcomes are measured longitudinally. Researchers often do separate analysis using these data with well-established statistical methods, including mixed effects models for repeated measures, and proportional hazards models for event data (Cox, 1972). However, the use of separate models may not be appropriate when the longitudinal variables are correlated with the participants health status and primary endpoints. Namely, when the longitudinal outcomes are of interest, medical events causing nonrandom missing or dropout is critical for the accuracy of the longitudinal analysis; when the analysis focuses on the risks and the survival time, without or simply adding the longitudinal responses which are usually error-prone as the time-varying covariates in the survival model may cause biased estimation. Nowadays, considerable scientific interest has focused on the relationship between the repeatedly measured variables and the distribution of the random event time. A method so-called joint modelling approach, in which the models of the event time distribution and longitudinal measures are conditionally independent based on a set of latent processes, has attracted increasing attention over the last two decades. The idea of coupling a survival model with a longitudinal model makes it possible to exploit the information contained in both data by dealing with the measurement errors and records that we do not have for the complete history of the past values.

The early work on the joint model, postulated by Faucett et al., 1996 and Wulfsohn et al., 1997, was primarily motivated from HIV studies, in particular, linking the trajectory of CD4 counts and time to death. Nowadays, joint modeling methodology has been widely used

and enriched. The typical setting of joint model for longitudinal and survival components consists of three parts: (i) a model for time to events, (ii) a model for the longitudinal markers trajectories, and (iii) linking both models through a latent structure underlying both (i) and (ii) processes (Proust-Lima et al., 2014). The key assumption and building block of this approach are that the two types of outcomes are conditionally independent given the common latent subject-level process. When multiple longitudinal measures are available, appropriate accounting for the correlations among these longitudinal outcomes can improve the joint model's efficiency and also the predictive ability. So far, two major approaches to link the component processes have been developed, considering either shared parameters or latent classes to capture the relationship between the longitudinal markers and the time to events (Proust-Lima et al., 2014). The preference for one construction over another depends on the real data structure, outcomes of interest and prior specifications. Both Bayesian/Markov Chain Monte Carlo (MCMC) and maximum-likelihood estimation (MLE) approach are used for the joint modeling analysis (Yang et al., 2016). Together with the theoretical development of the joint models, many powerful and flexible statistical software (e.g. SAS, R and WinBUGS) for conducting the joint modeling analysis are now conveniently available.

In general, the use of the joint modelling approach lets us to get a deeper insight in the relationship between the two analyzed processes. This method provides more accurate and insightful inferences for these repeatedly measured outcomes and offers prospective prediction of patients' risk to diseases with the adjustment of longitudinal information.

1.2 Dementia and Alzheimer's Disease

Before we get into the detailed statistical methods, we like to introduce and outline the main clinical application that guided the statistical development in this work.

Dementia is a general term for a group of brain disorders severe enough to interfere with person's daily life and now is a widespread and growing public health problem. There are a number of different types of dementia, but the primary ones are Alzheimer's disease (AD), vascular dementia, and Lewy body dementia (Zhang et al., 2017). As the leading type of dementia, AD accounts for some 60-80% of the dementia cases. In 2018, more than five million people in the United States were affected by AD and AD was reported as the sixth leading cause of death. About one in three seniors dies with AD or another form of dementia (Alzheimer's Association, 2019). Currently, there is no effective treatment to delay the onset or to slow the progression of AD.

1.2.1 Alzheimer's Disease Symptom and Progression

The first early sign of AD is change in memory, mainly with difficulty remembering recently learned information. As the disease progresses, it leads to severe symptoms including mood and behavior changes; deepening confusion about time, places and events; unfounded suspicions about family members and friends; difficulties with communication and language; and more serious memory loss and behavior change (Alzheimer's Association, 2019).

Like many other chronic diseases, AD is characterized by a very long multi-stage degradation process of cognition and large inter-individual variability, where preclinical symptoms could gradually worsen over a number of years. Figure 1 shows the stages from preclinical to AD onset: Stage 1 (Preclinical), no noticeable symptoms, but some identifiable biomarker changes in the brain; Stage 2 (Mild Cognitive Impairment), early symptoms of possible AD with measurable memory impairment and cognitive decline, but no large impact on daily life. It is an intermediate condition between normal cognitive aging and AD. However, MCI does not always lead to AD; Stage 3 (Dementia due to AD), significant cognitive decline, brain impairment, and major impact on daily life (Kumar et al., 2020). Knowledge of the

pre-diagnosis phase is very important for the understanding of the degradation process and the early detection of subjects at high risk of AD.

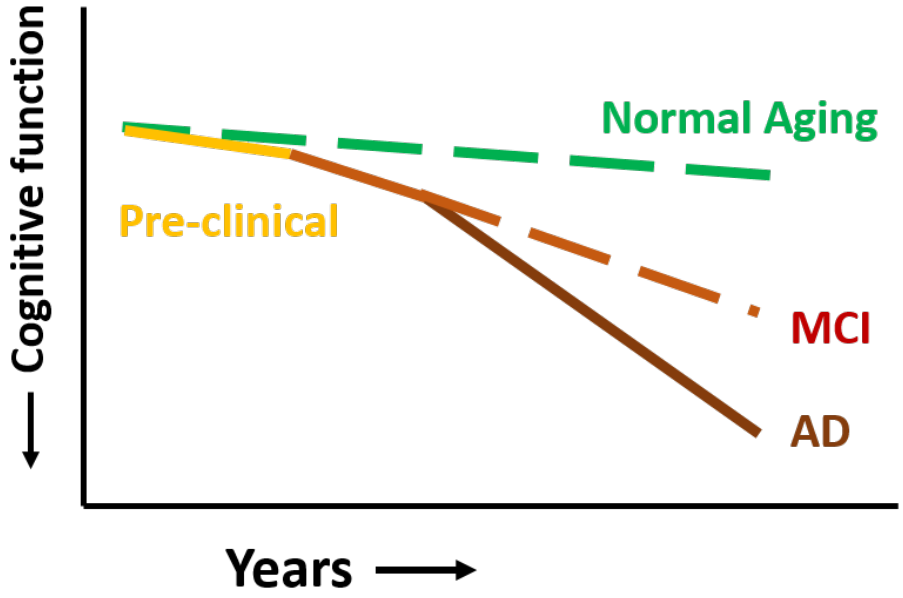

Figure 1: Alzheimer's disease progression with long degradation process of cognitive decline

1.2.2 Alzheimer's Disease Diagnosis

Onset of AD, time-to-event of interest in our study, is based on clinic AD diagnoses with standard criteria. Neurologists or trained professionals determine whether a person with memory problems is having "possible AD" where dementia may be due to other cause, or "probable AD" when no other cause for dementia can be found. The consensus diagnosis is based on neurological examination and medical history, cognitive performance testing. Clinical diagnosis of AD is different from neuropathological diagnosis of AD, where AD is definitively diagnosed after death, by examination of brain tissue in an autopsy (National

Institute on Aging, 2017). The onset of AD in most survival models is based on the clinical diagnosis.

1.2.3 Cognitive Decline and Cognitive Function Tests

It has been known that faster cognitive decline begins years before individuals develop dementia (Wilson et al., 2013). The early signs including having trouble remembering, learning new things, concentrating that affect person's daily activities.

To identify early signs of cognitive decline, the cognitive function assessments are used as a tool to assess the level of cognitive impairment and potential risk of AD. There are more than thirty different types of cognitive screening tests identified to indicate the likelihood of genuine cognitive impairment. Based on established neuropsychological profiles, it is suggested that a comprehensive screening instrument for the tests should include six core domains or abilities: "Attention/working memory, new verbal learning and recall, expressive language, visual construction, executive function and abstract reasoning." (Cullen et al., 2007). During a cognitive test, participants answer a series of questions and/or perform simple tasks. For example, the tester is asked to memorize a few words, identify a picture of a shape or object. The tests need no preparation from testers and are not recommended to be retaken avoiding learning effects. The most common types of test are Mini-Mental State Exam (MMSE) and Montreal Cognitive Assessment (MoCA) test. A summary score is provided with different range and cut-offs based on the type of the test. Usually higher scores reflect better cognitive function. A good screening test will have a high level of accuracy in differentiating normal cognition, cognitive impairment or cognitive disorder.

1.2.4 Alzheimer's Disease Research Cohorts

Since approximately 50 million people are currently living with AD or other types of dementia and cognitive decline is the hallmark feature of dementia, prevention of such decline is a public health priority. There is a worldwide effort underway to predict, delay, and treat this disease. Several Population-based cohort studies were designed and built during the last half century (Chibnik et al., 2017). These cohorts use consistent methodology throughout the course of data collection and provide needed information for understanding the epidemiology of AD, allowing for exploration of effect modifiers or interventions evaluation.

For example, the PAQUID (Personne Agee Quid) is a population-based cohort of 3777 elderly individuals recruited in 1988 in the southwest of France (Dartigues et al., 1991). There have been nine waves of data collection at different years after the initial data collection. Alzheimer's Disease Neuro-imaging Initiative (ADNI) is a global research study running since 2004, in which researchers at 63 sites across North America track the AD progression in the human brain with imaging, genetics and bio-specimen bio-markers through the process of normal aging, early and late MCI to dementia or AD (Petersen et al., 2010). Rush Memory and Aging Project (MAP) is an ongoing longitudinal and clinical-pathological cohort study of aging and AD, recruiting persons from across northeastern Illinois. Through July, 2019 the study has enrolled 2,134 individuals. Evaluations are annual and besides cognitive function, MAP study also supports studies of decline in motor function and disability (Bennett et al., 2018). The data generated in the MAP cohort are used in analysis presented in this dissertation and will be described in later sections.

1.2.5 Risk Factors and Treatment for Alzheimer's Disease Research

The exact causes of AD remain unknown. At one time, AD was thought to be genetic. Over the last thirty to forty years, researchers have identified many factors that may contribute to or prevent the disease. So far, research has linked the disease with the following potentially risk factors for AD: age (the strongest risk factor), gender, family history, genetic factors (heredity), head injury, cardiometabolic risk factors, life-style and activities and others (alcohol, depression, medication, dietary pattern etc.) (Alzheimer's Association, 2019). While some risk factors like age, genotype and family history, cannot be changed, some modifiable factors are involved in the AD progression. One promising line of research suggests that intensive risk factor modification, especially during midlife (age 45 to 65 years), has the potential to delay or prevent a substantial number of dementia cases worldwide (Press and Alexander, 2019). Prevention of AD by targeting risk factors including participating in regular physical activity, staying socially active, eating a healthy diet and maintaining good heart health, may decrease AD occurrence.

At this time, there is no cure for AD, and the treatments are largely ineffective. However, treatment for symptoms are available, and these drugs (donepezil, rivastigmine, galantamine and memantine) for this disease do not decelerate or prevent the progression of the disease (Tan et al., 2014). Most medicines work the best for patients in the early or middle stage of AD. This means that it is possible to delay the symptoms of dementia and improve quality of life in later life by taking control of the health in young and middle-aged adult years. Indeed, subjects with early detection with high risk of AD could be the target population for new treatments. In ongoing clinical trials, many possible interventions are developed and tested, including cognitive training, drug therapies, and controlling risk factors that can be managed with effective treatments and/or through healthy lifestyle choices.

1.3 Statistical Methods for Alzheimer's Disease study

1.3.1 Time-to-Event Analysis of Alzheimer's Disease

To identify potential risk factors causally associated with AD onset, many studies applied proportional hazards regression model (Cox, 1972; Kukull et al., 2002), which is a commonly used statistical method in biomedical research for investigating the association between the time to clinical event and one or more predictor variables. Survival time is calculated either from study entry or from retrospectively estimated dates of disease onset. A key assumption of this model is that the ratio of the hazards for any two groups is constant over time, thus the hazard curves between groups should be parallel and cannot cross.

Demographic information (e.g. age, sex, education level, and Apolipoprotein E (ApoE4) genotype) and major medical comorbidities (e.g. heart disease, hypertension, stroke, and diabetes) up to AD diagnosis are usually considered as covariates in AD related survival models (Helzner et al., 2008).

Although AD is a common disease in elderly, some people do not develop AD through their life time and are free of AD at death. An intuitive consideration is that those who develop AD in late life are those who lived long enough to enter the age of risk for developing AD. Because the incidence of dementia rises sharply at ages greater than 75, many population-based cohort studies and clinical trials on AD focus on those people aged 65 years and older. Recruitment of an older sample is thus inevitably biased in favor of survivors (Chang et al., 2012), so called survival bias. When death is considered as the main reason for the dropout in study cohort versus other causes, the true relationship between risk factors with AD incidence may not be identified using conventional survival models that treat mortality as uninformative censoring. Death occurs before AD onset, and this phenomenon is referred to as competing risks (Hernan et al., 2008). Like many other statis-

tical approaches, the competing risk analysis includes both parametric and non-parametric methods. Identifying and addressing competing risks will help eliminate or reduce bias in predicting the probability of developing AD. Besides the cause-specific hazard (Satagopan et al., 2004), another model called subdistribution hazard model proposed by Fine and Gray, 1999 allows to direct assess the effect of the target factor on the marginal probability function, regardless of whether the subjects are censored or failed from other competing events (Kleinbaum et al., 2012).

Another challenge is that AD can only be diagnosed during the clinic visit, but the exact time when AD is developed should be between the visit of diagnosis and the visit before it, resulting in interval-censored data where time of AD occurrence can not be observed exactly. Moreover, the aging data usually requires a long time to collect and clinic visits depend on the availability of the participants and financial support of the study. This often causes some gaps in time between designed AD screenings. Regarding these issues, many methods that can handle interval-censored survival data have been developed and applied in the AD research area (Rouanet et al., 2016; Yu and Ghosh, 2010).

1.3.2 Longitudinal Analysis of Cognitive Decline

The longitudinal designs for AD studies usually extend over a long period of time, consisting of several follow-ups at fixed visit time points. The visits are often years apart to avoid any learning effect in the tests. During each AD screening, cognitive function test scores as well as a broad range of other health or aging related factors are collected. These longitudinal data motivate researchers to apply methodological tools to take insights into age-related degradation of neuronal functions and the factors contributing to the degradation and AD.

Cognitive function test results collected as repeated measurements are usually discrete quantitative outcomes consisting in summary scores, which are usually continuous and roughly normally distributed. To describe the change over time of the cognition and its association with predictor variables, numerous exploratory analysis were performed using linear mixed-effects model (LMEM) during the past decade (Lundervold et al., 2014; Sabia et al., 2017). As a method for analyzing data from repeated measures designs, LMEM has increased in popularity in the last decade. By definition, mixed models involve both fixed and random variables and the random effects structure provides more accurate estimation by aiding correct inference about fixed effects (Harrison et al., 2018). The LMEM relies on three assumptions that first, the outcome is continuous, second, the random components are Gaussian, and third, change in the explanatory variable is associated with constant fixed change in the outcome (Proust-Lima et al., 2011). The introduction of mixed fixed and random structure affords several non-exclusive advantages over more traditional techniques like repeated measures ANOVA which is antiquated. Another benefit is that mixed models do not require a complete data set when missing data are assumed to be random.

To study the cognitive function change across years, the typical setting for random effects is intercept and slope. The covariates include age, gender, education level and ApeoE genotype (Hester et al., 2005). The interaction between factor and time of enrollment is used to test for the differences in the effects on the rate of cognitive decline between the categories of factor. If a linear trend cannot be assumed and a non-linear decreasing trajectory is considered, quadratic time term, B-splines or change time points are usually used inside the linear mixed models.

1.3.3 Joint Modeling of Cognitive Decline and Alzheimer's Disease

The cognitive decline and the AD incidence should be analyzed simultaneously as they are highly correlated. Taking advantage of the repeated measurements of cognitive performance and AD clinical diagnosis information both collected in a longitudinal structure, a joint modeling approach can be applied. On one hand, cognitive impairment is a predictor of the AD incidence; on the other hand, the follow-up is often truncated by AD, accounting for the non-random dropout of patients and missing data. Model jointly analyzing cognitive decline and incident AD is considered as a powerful statistical tool capable of capturing the association between the two processes.

This approach has been extended in different ways. AD dementia associated demographics, genomics, lifestyle behaviours, psychometric markers, vascular factors, neurological and neuropsychological factors are also monitored during the studies in most aging projects. An interesting extension would be to model multiple longitudinal outcomes with AD incidence. For instance, it has been reported that motor impairments are prior to cognitive impairment and changes in motor function can be consider as potential clinical responses for early detection of dementia (Kueper et al., 2017). From the clinical point of view, investigating multiple correlated measurements of biomarkers jointly could provide a better view of disease progression process and a better prediction on the risk for AD conversion than a single clinical measurement. It is also interesting to address if the association may be affected by a host of factors including age, genotype, gender which are assumed not to change throughout the study cohort.

Furthermore, the joint modeling approach can further incorporate competing risks from the terminal event, such as death, a changepoint in the trajectory for a particular functional decline, functional data analysis through a functional joint model (FJM), and dynamic prediction framework for predicting the trajectories in future timeline and risk of clinic event

(Dantan et al., 2011; Li and Luo, 2019). More details on these extensions are provided in Chapter 2.

1.4 Outline

This dissertation aims to appropriately analyze longitudinal and time-to-event data jointly, explore the associations and describe the trajectories and characteristics of both processes. In the next chapter, we provide a detailed review of literature on the development of a joint modelling framework and the methods which have been applied on the joint modeling of cognitive aging and risk of AD. We discuss about the methodological challenges raised by taking into account the order that existed in the degradation process, the heterogeneity between individuals of the decline trajectory and informative dropout. We also describe details of the Rush Memory and Aging Project cohort which is used in this study. In the third chapter, we propose a Bayesian joint modeling analysis to link the bivariate longitudinal outcomes with the risk of an event under latent class framework and we also show the alternative joint modelling method using shared random effects structure. In the fourth chapter, we introduce a joint model of bivariate random changepoints in the longitudinal process coupled with competing events. The last chapter is the application of our methodology to MAP data.

2 LITERATURE REVIEW ON JOINT MODELING OF COGNITIVE DECLINE AND ALZHEIMER'S DISEASE AND MAP COHORT

In clinical and epidemiological research studies, repeated measurements of clinical variables, event history, and other types of data are simultaneously collected. Depending on the study interest, different types of statistical analysis methods are performed.

The clinical studies are often designed to address the primary question on the effect of an intervention or a risk factor on the survival probability. When investigating the relationship of survival time to longitudinally collected clinic markers, survival models, such as proportional hazard (PH) model or accelerated failure time (AFT) model with time-varying covariates, are considered as a powerful tool. However, in most cases, the observed repeated measured covariates may not be the true values and contain intra-subject errors. For example, cognitive function, which is not directly observed but measured through numerous psychometric tests, contains a great deal of random errors between and within subjects. Replacement of the true value with error-prone time-varying covariates in the survival model may cause biased estimation. Moreover, the time-varying covariates values are only known at the time points at which they are measured and thus are assumed to be constant between measurements in the model. In addition, missing values on covariates are common, and additional steps such as multiple imputation are required. On the other side, when the aim is to study the longitudinal trajectories, for example, cognitive decline, the follow-ups of longitudinal data may be truncated by the dropouts due to different reasons. The occurrence of the clinic event may induce non-random missing data. For example, subjects with more severe pre-AD phenomenon may be more likely to withdraw from the study than the healthier participants. There is no statistical test to check data missing at random (MAR). Failure to take in count of this trend, ordinary longitudinal analysis like LMEM may result in biased estimation.

The concern of biased inferences leads to considerable recent interest in a joint model to link the longitudinal and time-to-event data. In the last decade, in order to highlight the multi-factorial nature of the marker-event relationship, for example, between AD and cognition trajectory, many methods have been proposed to incorporate the two processes. The statistical frameworks provided are beneficial and give us directions for future analysis.

2.1 Joint Model Frameworks

2.1.1 Shared Random Effects Model

The early development of the joint model was primarily motivated from HIV/AIDS studies, in particular, linking the trajectory of CD4 counts and time-to-death (Ibrahim et al., 2010). Many of the works on joint model in early 1990s focus on imputing appropriate value for longitudinal variable to reduce the bias due to failure to account for measurement error (Self and Pawitan, 1992; Raboud et al., 1993; Tsiatis et al., 1995).

Rather than relying on approximations, Henderson et al., 2000 proposed a flexible joint model in which survival and longitudinal data are conditionally independent given a zero-mean multivariate latent Gaussian process, so called shared random effects model (SREM). This design is based on very strong assumptions about the association between the two outcomes and a set of random effects is assumed to capture their interdependence. Shared random effects model belongs to the class of shared parameter models, which usually rely on a structure of homogeneous sub-population and the trajectories of the markers smoothly distributed and linked to the risk of event. An excellent general review article on shared random effects models was given by Tsiatis and Davidian, 2004. In the studies of cognitive function and risk of dementia, this shared random effects framework has been largely applied (Gao, 2004; Li et al., 2017).

Shared random effects approach has been applied for the joint analysis of longitudinal data with binary outcomes (Horrocks and Heuvel, 2009) or multivariate normally distributed longitudinal biomarkers (Rizopoulos and Ghosh, 2011). These models were soon incorporated with multiple recurrent or competing events, and therefore needed more complex association structures. A very recent review paper of Papageorgiou et al., 2019 focused on the shared parameter formulation of the joint modeling incorporating multiple longitudinal outcomes of varying types.

2.1.2 Latent Class Model

Lin et al., 2002 considered a related model called joint latent class model (JLCM) in which the common shared random effects are replaced by a latent class framework that accommodates underlying population heterogeneity to describe the differences in prostate-specific antigen trajectories and prostate cancer. Compared to the shared random effects joint model, JLCM has limited literature, but is considered to be particularly suited for heterogeneous data. In 2003, Hashemi and Jacqmin-Gadda applied the latent class joint modeling method on cognitive decline and AD in the PAQUID cohort (Hashemi and Jacqmin-Gadda, 2003). This was one of the first works bringing the joint modeling approach to the field of AD research. The JLCM assumes that the correlation between the longitudinal data and event outcomes is fully captured by the latent classes. This model stratifies the heterogeneity in marker trajectory and the time to an event through latent classes, assuming cases in each subgroups sharing the same longitudinal marker trajectory and the same risk of the event. For instance, in dementia applications, to distinguish different homogeneous sub-groups, subjects are allocated to different latent classes corresponding to cognitive ability and risk of AD with the highest posterior probability, with the adjustment of covariates. Each class has class specific transition intensities to AD and cognition trajectory. Because of the flexibility in modeling of the dependence between the longitudinal marker and time to event, as well

as the covariate effects trajectories, extensions and innovations have been proposed afterwards. Proust-Lima et al., 2014(2) extended the JLCM to analyse simultaneously multiple longitudinal markers from cognitive data and risk of dementia.

One of the drawbacks of the JLCM is that as the sub-models are class-specific, assumptions of proportionality between classes need to be made. Otherwise, too many parameters are involved especially for large number of latent classes. In addition, when the number of classes is too small or too big, the model has a low discriminatory ability and interpreting the latent classes can be difficult.

Both SREM and JLCM rely on the conditional independence assumption of survival and longitudinal processes. But the bases of the assumption of the two frameworks are different regarding the link between the longitudinal and event onset processes. A homogeneous population is assumed in SREM, in which the trajectory of the longitudinal outcome is the same and the relationship between the marker and event risk is continuous. In contrast, JLCM accounts for the heterogeneity of the population with latent class categorical structure, where each category is under a class-specific average profile of the longitudinal marker and risk of event. Thus classification is essential. The random effects in the JLCM are endogenous in the pathway of latent class (see Figure 2). Proust-Lima et al., 2014 provided a good review of latent class methods and strategies.

2.1.3 Copula Approach

Another approach to the challenge of linking these two types of outcomes is using a copula model where Gaussian copula function constructs a joint framework by combining the two marginal distributions relying on a dependence structure (Rizopoulos et al., 2008; Ganjalia and Baghfalaki, 2015). The parameter estimation is conducted with the marginal distribution and specified copula function. This method is considered as a tool to construct

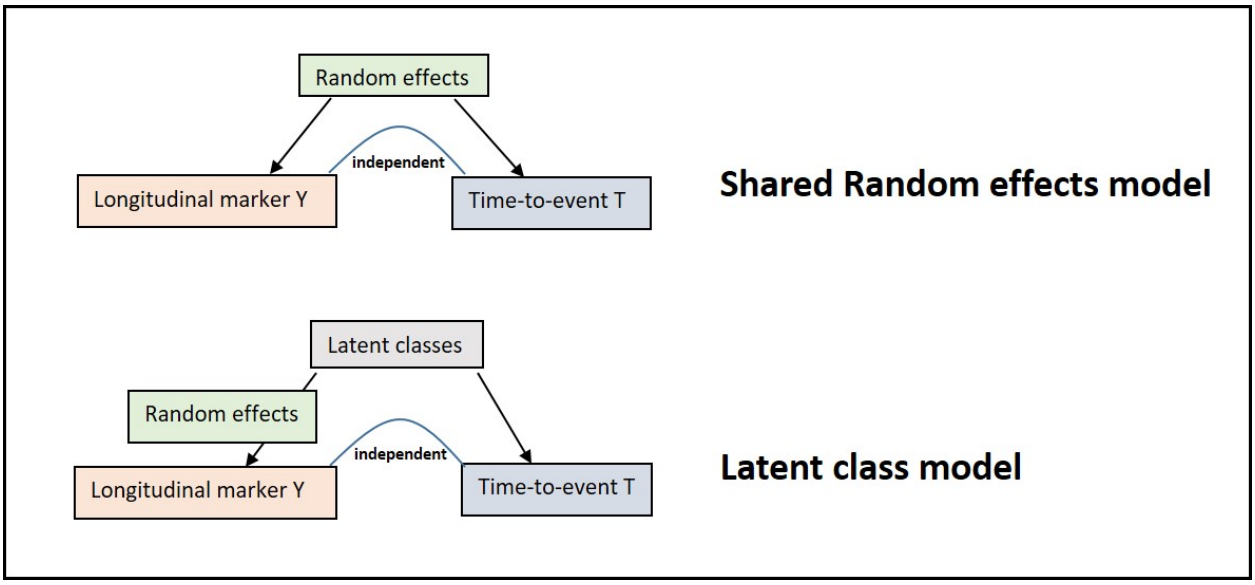

Figure 2: Shared random effects model vs Joint latent class model

joint density for an alternative reparameterization for the shared random effects model. The copula approach has not yet been applied to the study of cognitive decline and AD.

2.2 Parameter Estimation Methods in Joint Models

2.2.1 Maximum Likelihood Estimation

Maximum likelihood (ML) method is commonly used in joint models. Maximum likelihood estimation (MLE) finds the values of the parameters by maximizing the likelihood of the model and provides unbiased estimates under the true random effects distribution assumption.

The MLE approach has been applied extensively in joint models in AD research, especially in JLCM. Hashemi and Jacqmin-Gadda, 2003 performed maximization of a pseudo-likelihood using Newton-Raphson type algorithm on a latent process model for jointly modeling demen-

tia and cognition. Later, the Marquardt algorithm for optimization, a robust Newton-like algorithm, was applied to find the MLE in joint model with multivariate cognitive measures and competing risks of dementia and death or interval-censored dementia (Proust-Lima et al., 2014(2); Rouanet et al., 2016). The expectation-maximization (EM) algorithm was also implemented in the joint model to deal with the unobserved random effects. The EM algorithm offers many computational advantages over direct likelihood maximization and is commonly used for complex likelihood functions. Li et al., 2017 estimated the parameters of functional joint model (FJM) via EM algorithm to account for functional predictors on AD progression. Akaike information criterion (AIC) and Bayesian information criterion (BIC) are common methods for model scoring and selection under the MLE framework. However, data with substantial measurement error or sparsity may decrease the reliability of estimates. Also computation of multi-fold integration can become intractable when the mixed effects model is dealing with multiple random effects.

Maximum likelihood estimation of joint model can be implemented using the R package "JM" (Rizopoulos, 2010), in which the log-likelihood maximization is using a hybrid optimization procedure in which the estimation is first through EM algorithm, then switched to use a quasi-Newton algorithm if convergence is not reached. Another R package "lcmm" focuses on JLCM and maximum likelihood estimators are obtained based on a modified Marquardt algorithm (Proust-Lima et al., 2017). As statistical software is updated regularly, many additional options are increasingly available including specification of the association structure, competing risk settings, predictions, and plots etc.

2.2.2 Bayesian Markov Chain Monte Carlo Method

The Bayesian approach nowadays has been successfully used for parameter estimations for different complex modeling situations in various scientific research. The fast development of

Bayesian statistics has been seen in the past two decades. In Bayesian approach, parameters are treated as random variables and this is the major difference from the classical statistical theory. The distributions of the parameter are called priors. The likelihood of observations is effectively weighted by the prior distributions. The resulting posterior distribution is the essential outcome used to obtain the estimates of interested parameters. The Markov chain Monte Carlo (MCMC) method, especially the Gibbs sampler, is a type of algorithm now mostly used for parameter estimation from models with hierarchical structure, and can be simpler to implement than the direct maximization of likelihood function. Compared to ML method, which needs direct high dimensional integration, the MCMC sampling provides numerous solutions to problems by samplings from posterior distribution and computing each parameter distribution. Thus, the parameter estimation is much straightforward under the Bayesian paradigm through the posterior distribution. For model evaluation, DIC (deviance information criterion) and WAIC (widely applicable information criterion) have been widely used to compare between Bayesian models.

Regarding the computational complexity for joint models where the dimension of the random effects is large, Bayesian approach can provide alleviation of the computational burdens. Indeed, many developments of joint models have been made in a Bayesian framework recently (Lawrence et al., 2015; Dessiso and Goshu, 2017). In AD dementia and cognition research, the Bayesian method was first considered by Hall et al. for parameter estimation from random changepoint models of the cognitive decline in dementia population (Hall et al., 2003). Later on, many AD studies have adopted Bayesian procedure due to its ability to deal with multiple longitudinal outcomes and competing risks in joint models (Yu and Ghosh, 2010). It is known that, in AD research, in order to closely monitor cognition, multiple psychormetric tests are ordered. These data provide a complex longitudinal framework that needs to appropriately account for potential correlations. Moreover, as the risk of competing events increases with age, a more complex event structure is required in the survival submodel. Estimation in high-dimensional joint models where likelihood is far too complex to

be computed, is feasible by sampling from it. Another important application of the Bayesian framework in AD research is to predict the efficacy and safety responses of future patients conditional on the data observed so far. Li and Luo, 2019 developed a Bayesian functional joint model to obtain accurate inference and provide better prediction for new subjects by incorporating many features in longitudinal and survival data.

As the Bayesian MCMC method has the big advantage of dealing with computational complexity, one thing we should be cautious about is that the results may rely on the prior specification. The prior distribution can incorporate additional information from similar studies within a joint probability model. At the same time, the choice of the prior should be decided with care. In addition, in complex models with many parameters, autocorrelation and convergence needs to be monitored in MCMC approach. The convergence of the MCMC sampler after burn-in can be determined by several criteria. Gelman-Rubin diagnostic test (Gelman and Rubin, 1992) and Geweke test (Geweke, 1992) are commonly used. Geweke test checks the convergence based on a test for the difference between the mean of the first part and last part of a single chain. The Gelman-Rubin diagnostics test calculates the "potential scale reduction factor" for each parameter. A general guideline suggests that a values less than 1.1 is good. One drawback of Bayesian approach is that each iteration of the algorithm requires accessing the whole data which in practice makes MCMC too slow to reach convergence especially when the dimension is large. Replacing the exact posterior with partial posterior marginals is a possible solution.

Bayesian inference gains wide popularity with the availability of ready-to-use software packages developed for Bayesian analysis based on MCMC. Many are free to download. WinBUGS (Lunn et al., 2000) developed by the BUGS (Bayesian inference Using Gibbs Sampling) project team is a powerful and flexible statistical program to perform Bayesian analysis. In WinBUGS, the user only needs to provide the data and initial values and specify the structural of the statistical model (prior and likelihood). The software will then carry

out one of the MCMC algorithms to generate samples from the posterior distribution of the specified model. Another software, OpenBUGS (Lunn et al., 2009) released in 2005 can use more than one algorithm for the simulation for the class of full conditional distribution of each node, providing a more flexible and extensible operating environment. For joint model analysis, the R package "JMbayes" (Rizopoulos, 2014) fits the models for longitudinal and time-to-event data in a Bayesian framework and several types of association structure between the two outcomes are provided.

2.2.3 Two-Stage Method

Two-stage method was developed by Self and Pawitan, 1992. Although it is not as popular compared to MLE and Bayesian methods, it has been used in parameter estimation under joint longitudinal-survival framework. In this approach, parameter estimation is conducted separately in two-stage procedures. The first stage consists of modeling the longitudinal components through mixed effects model without consideration of the survival information, and in the second stage, the modeled values from the previous step are included as time-varying covariates in a time-dependent survival model.

Two-stage method gains some advantage in ease of use and fast computation. However, this is not an unbiased approach. When determining the correct longitudinal estimates, it ignores the survival information, and so, possibly causes selection bias due to failing to account for informative drop outs, which can result in uncertainty in modeling the longitudinal process. This was discussed in detail in a paper of McCrink et al., 2003.

2.3 Extension of Joint Model and Methodological Challenges in Alzheimer's Disease

2.3.1 Joint Model with Non-Linear Longitudinal Outcomes

Several studies have shown that the slope of global cognitive decline was gradual at first and then decreased dramatically within a few years prior to the diagnosis of AD or before death (Wilson et al., 2007; Sliwinski et al., 2006). Also, there should be a wide individual difference in rates of change in cognitive function as AD progresses, especially during the late-life period (Amieva et al., 2008; Petersen et al., 2001). Moreover, it is believed that the acceleration of the cognitive decline manifests through a changepoint that may depend on subject-specific characteristics. The time point at which cognition evolution of the patients who will develop AD in future become distinguishable from normal aging people is essentially important (Amieva et al., 2014; Jansen et al., 2018). As more effective treatments become available for AD, this time frame will provide invaluable information for developing new treatments and designing prevention strategies at earlier stages of AD.

Regarding the non-linear trajectory of cognitive function, some studies modeled the degradation process using quadratic time term, however, this setting does not allow identification of the moment when the change in rate takes place. In 2000, Hall et al. introduced a changepoint in cognitive decline as a parameter in the piecewise linear model (Hall et al., 2000). The changepoint is defined as "an estimated time before diagnosis of AD, at which the rates of decline among vases and non-cases begin to diverge". Changepoint mixed modeling has been used to describe the trajectory of a longitudinal measurement and detect the change in the trend since the 1970s (Hinkley, 1970; Smith, 1975; Carlin et al., 1992). Hall compared the rate of decline among those who developed AD to those who remained free of AD. Profile likelihood method was used and this method required a common changepoint for all subjects who had a changepoint. However, since AD dementia is a heterogeneous disorder

with multiple phenotypes and genotypes, this same changepoint assumption might not be appropriate. Hall later proposed a random changepoint model in which the time point of change in decline rate was subject specific through a Bayesian MCMC approach (Hall et al., 2003). In 2006, Jacqmin-Gadda et al. pointed out that the estimates of changepoint only using the data from subjects diagnosed as demented during the follow-up contained selection bias (Hall et al., 2003). Patients without AD at the end of the follow-up might be in the preclinical phase and treating them as nondemented is not appropriate. Subsequently, to avoid the selection bias, Jacqmin-Gadda combined a piecewise mixed model with a random changepoint for cognitive decline and a survival model for AD risk. The changepoint was used to link the two parts (Jacqmin-Gadda et al., 2006). A direct-likelihood approach was used for parameter estimation.

Testing the existence of a time point of change in rate in a mixed model with repeated measures is an area of interest in biomedical and epidemiological studies, especially in respect to the natural history of chronic diseases. Very recently, motivated by the study of AD dementia, Segalas et al., 2019 described a novel way to test the existence of a random changepoint in a mixed model. This test was applied to study the shape of the pre-diagnosis cognitive decline among the elderly and smooth transition between the two linear phases of cognitive decline was considered in the model.

In above research, the change in rate of the cognitive decline was described with an abrupt transition in a piecewise linear model, which is simple and offers the advantage of detecting a significant immediate departure in the direction. However, this setting is artificial and cannot not always be appropriate and realistic in practice. Indeed, given the nature of cognitive decline, the entire trajectories are generally smooth, even at the changepoint. Alternatively, several types of smooth changepoint models were proposed to imply a graduate change between the two linear phases. Based on how the trend of transition is quantified, changepoinit models can be broadly classified into two families: (i) piecewise (broken-stick)

model, (ii) smooth changepoint model (Bacon-Watts model (Bacon and Watts, 1971), bent-cable (Chiu et al., 2006), polynomial regression model (Hout et al., 2010). In application to data of cognitive aging, Hout et al. introduced the smooth random changepoint in modelling of cognitive decline and both polynomia and Bacon-Watts functions were applied in a mixed-effects model. Later, Yang and Gao, 2013 compared the Bacon-Watts model with the polynomial regression model and proposed a bivariate random changepoint model to jointly model cognitive function and body mass index (BMI) over age. Both groups showed good performances of the Bayesian approach for parameter estimation in the complex model structure. One limitation is that only the subjects who were diagnosed as demented during the follow-up were selected in these analysis of smooth random change. This may lead to a selection bias due to dropouts caused by death or severe pre-dementia symptoms. Also subjects considered as dementia-immune during the study and diagnosed after the end of the study were excluded. To overcome this limit, an extension as a joint modelling approach with adding AD dementia-death information would be more appropriate.

2.3.2 Joint Model with Competing Risk and Multistate

In the spirit of Hall and Jacqmin-Gadda's work, Yu and Ghosh, 2010 proposed a joint model that accounted for a changepoint for cognitive trajectory and mixture survival time for dementia onset and death. As stated before, death is a major competing risk for AD dementia. For a better estimation for the effect of covariates on AD related outcomes, the dropouts as death cannot be ignored. The cause-specific hazard model for competing risks is commonly used. This model included three types of status in the survival part: (i) still alive and dementia free, (ii) dementia developed, and (iii) deceased. Yu further divided the alive or deceased into: immune to AD or having the potential to develop AD. The parameters in this complex model were estimated by MCMC method, which made the computation feasible. Another competing risk analytic method called subdistribution

hazard model proposed by Fine and Gray, 1999 is also frequently applied in AD dementia survival analysis (Li, 2016; Kuo Hout et al., 2019). This model allows direct assessment of the effect of the target factor on the marginal probability function. Marginal probability is defined as the probability of onset of a particular event, regardless of whether the subjects are censored or failed from other competing events (Kleinbaum and Klein, 2012). Like many other statistical approaches, both above competing risk models include parametric and non-parametric methods. It is believed that identifying and addressing competing risk will help to eliminate or reduce bias in predicting the probability of developing AD.

As an useful improvement, considering the risk of dementia should be increased in the phase of accelerated cognitive decline, recently Dantan et al. combined multistate survival models and mixed models for longitudinal outcomes, assuming the risk of dementia is null before entering the accelerated decline phase in cognition (Dantan et al., 2011). The multistate model focused on 4 states (healthy, pre-diagnosis, dementia and death) and the transitions between them. This new model is viewed as an improvement on the random changepoint model by handling of informative right censoring due to death.

2.3.3 Other Extensions

Mental health disorders in seniors are very complicated and look different in everyone. Statistical methods are being developed and extended to overcome the limits of previous models and allow us to consider more complex conditions and connections in the real world.

Besides what have been mentioned above, there are some other extensions that have been applied on the survival analysis. First, aging study, with a long-running cohort, could have big gaps in time between visits. The AD status is evaluated intermittently as cases can only be determined during clinic visits. To take into account the uncertainty on the time of AD dementia, Rouanet et al., recently proposed a joint model for interval-censored events and

cognitive decline (Rouanet et al., 2016). Secondly, traditional AD risk analysis assumes that everyone will eventually develop dementia and this assumption was challenged by the work of Zhou, 2013, who introduced the concept of immune subgroup for AD research (Stong, 2013). His work hypothesizes that some people may be immune to AD. Using a logistic regression formula to model the immune probability, and a Weibull distribution to model the survival function for those at risk, they obtained highly statistically significant evidence for the existence of an immune subgroup (Zhou, 2013).

In the investigation of the longitudinal mark trajectories, we have seen most articles focusing on one type of longitudinal outcome in the joint model. As multiple outcomes are collected during the clinical research, it will be informative to investigate the changes over time in multiple correlated outcomes. For example, the cognition function is measured by a group of cognitive tests that are correlated but functionally different. Hall et al., 2000 proposed a bivariate changepoint model in which the changepoints in two different measurements of cognitive function over age were compared. Only the intercepts of the two scores were correlated in the model. Later, Yang and Gao, 2013, investigated the relationship between the changepoints of cognitive and BMI measurements in a bivariate changepoint model. In this work, only the correlation of time of changepoints was considered between the two markers. These correlation structures are too simple to describe the association between trajectories of the markers in the cognitive system. Very recently, Segalas et al., 2020 proposed a bivariate random changepoint model in which the associations between the marker-specific random effects (intercept, slope, and changepoint) were all considered.

While investigating the nature of the relationship between AD risk and cognition, it is important to understand the whole shape of the cognitive decline before and also after AD onset. Most of the analysis carried so far for joint modelling of cognitive decline and AD dementia are based on the observations before or at event onset. One may wonder how the slope of cognitive change after AD onset for those who still survive. Such extension

in the model may involve post-diagnosis survival time and two or multiple changepoints in cognition. Answers to these questions will be extremely helpful in early diagnosis of dementia and later investigation on the treatment effectiveness.

2.4 Rush Memory and Aging Project Cohort

The motivational application for this research is based on the Memory and Aging Project (MAP) of Rush University Medical Center. MAP study primarily enrolls senior residents in the Chicago area. MAP project is a longitudinal, epidemiologic clinical-pathological cohort study. Participants receive assessment of risk factors and detailed annual clinical evaluation for the common chronic conditions of aging with an emphasis on AD related outcome measures. This ongoing open cohort study began in 1997 and by July 2019 the study has enrolled 2,134 individuals with a mean age of 80.0 years, 14.9 years of education at baseline and 73.5% are female.

Sources of data include interviews, cognitive assessments, clinical evaluations, etc. These data have been carefully created and reviewed. The methods and resulting variables that were used in this work are listed below. The study design allows different types of analysis to be conducted for the investigation of related AD risk factors, AD incidence, decline in cognitive and other related outcomes.

2.4.1 Cognitive Function

A battery of 21 cognitive performance tests is administered annually for each participant in MAP. These tests assess a range of cognitive abilities including: Logical Memory I, Logical Memory II, Immediate story recall, Delayed story recall, Word List Memory, Word List Recall, Word List Recognition, Boston Naming Test, Category Fluency, National Adult

Reading test, Digit Span Forward, Digit Span Backward, Digit Ordering, Symbol Digit Modalities Test, Number Comparison, Stroop word reading, Stroop color naming, Judgment of Line Orientation, Standard Progressive Matrices (Bennett et al., 2012). Raw scores are converted to Z-scores and averaged to yield a summary measures of global cognitive function using the means from all participants.

2.4.2 Motor Function

The evaluation on motor function is based on the Unified Parkinson's Disease Rating Scale. Upper and lower extremity motor strength and performance tests are administered. These include: Grip and pinch strength measured by hydraulic dynamometers, arm abduction, arm flexion, arm extension, hip flexion, knee extension, plantar flexion, and ankle dorsiflexion measured with handheld dynamometry; time and number of steps to walk 2.4 meters and to turn 360°; participants are asked to stand on each leg and then on their toes for 10 seconds. The number of steps off the line is recorded when walking 8 feet heel-to-toe; and Purdue pegboard and finger tapping (Bennett et al., 2012). A global motor function summary score is created based on the z-score of each test result from all participants.

2.4.3 Alzheimer's Disease Diagnosis

Status of AD is evaluated at each MAP visit. The clinical diagnosis of AD was made by neurologists or trained professionals using data including: A structured neurological examination and medical history, cognitive performance testing, and with the assistance of an algorithmically based rating of cognitive impairment. The AD diagnosis was based on criteria of the joint working group of the National Institution of Neurological and Communicative Disorders and Stroke and the AD and Related Disorders Association (Morris et al., 2015).

2.5 Research Interests

We have explored the use of joint models in the literature focusing on the cognition and AD dementia research. The large number of different measurements and sufficient follow-up time that MAP provides allow us to carry out different statistical analysis to test our hypotheses for this thesis study. To our knowledge, the joint modeling of cognition trajectory and AD risk approach has not been yet performed in the MAP cohort study. We are interested to see how the joint modelling approach improves the fitting to the data. Shared random effects and latent class frameworks are both of interest to be applied. Bayesian estimation methods will be performed, MLE method can be used as reference. As far as we know, Bayesian inference has been barely used in the joint latent class model in AD research.

Secondly, it is widely known that cognitive decline and other factors such as BMI or motor function decline, coexist with aging. In particular, in the MAP cohort, the data support that decline in cognition and motor function may share a common causation and motor impairment could serve as a phenotypic marker of pre-clinical AD (Buchman and Bennett, 2011). Dr. David Bennett, the director of the RADC at Rush University Medical Center, mentioned: "Motor function is related to changes in cognition. There are a lot of data now showing that loss of cognition and motor function is due to a lifelong series of events." Instead of estimating the longitudinal models of cognition and motor separately, introducing the motor function into the joint model of cognition to have bivariate longitudinal outcomes would be more appropriate and this has never been investigated before. We like to propose a joint model with a more complex framework considering between markers correlation coupled with AD risk.

Thirdly, the smooth random changepoint models used to describe the two-phase of the longitudinal marker trajectory in previous studies did not take into account the health outcome data (mostly time-to-event) or was fitted to a particular subgroup, e.g. dementia cases.

To avoid selection bias, we like to extend the approach to a joint model for the changepoint and the time to event. In the model, we assume that the random changepoint is associated with some demographic covariates. The model will be enriched with competing events and interval censors to make it fit a more realistic situation. Additionally, as multivariate longitudinal responses can be attributed to underlying pre-diagnosis disease states, we like to further extend the joint model for bivariate longitudinal outcomes with random changepoint and time to event. We adopt the Bayesian procedure using MCMC sampling method for parameter estimation and inferences for these complex models. As an application, we like to fit the proposed joint model on the data in the MAP cohort. The investigation focuses on the correlation between the changepoints of cognition and motor function decline and comparison of the time of change given the risk of AD/death information. In the meanwhile, we will see how this random changepoint model for the markers of aging improves the prediction of the AD risk through shared random effects.

3 JOINT MODELING OF BIVARIATE LONGITUDINAL OUTCOMES WITH TIME-TO-EVENT DATA

3.1 Methodology

3.1.1 Mixed-Effects Models

The most popular framework for longitudinal data analysis is the mixed-effects model. This model allows each subject in the sample has subject-specific evolution across time. Let $Y_i(t_j)$ denote the follow-up measurements for i^{th} subject (i=1, 2, ... , n) at a specific time t_{ij}, j=1, 2, ... , n_i, assuming n subjects under study. The mixed model is written as

$$Y_i(t) = Y_i(t_{ij}) = x_{ij}^{\top}\beta + z_{ij}^{\top}b_i + \epsilon_{ij}, \tag{3.1}$$

in which

$$\begin{aligned} b_i &\sim N(0, \Sigma_b), \\ \epsilon_i &\sim N(0, \Sigma_\epsilon), \end{aligned} \tag{3.2}$$

in which β is a vector of the regression coefficients for the fixed effects $x_{ij}^{\top} = [x_i^{base}]^{\top}$, $[x_{ij}^{time}]^{\top}$, $[x_{ij}^{base\times time}]^{\top}$ that consist of time, covariates, and their interaction, respectively; $z_{ij}^{\top} = [z_{ij}^{time}]^{\top}$ denotes the row vector of the design matrix for the random effects b_i. Particularly, b_i are assumed following a multivariate normal distribution with zero-means and variance Σ_b which can be an unspecified matrix. In this way, the fixed and random terms measure the population-level effect and subject-level effect, respectively. The measurement errors ϵ_i are independent of b_i. If no autocorrelation is specified, these errors are assumed to be independently Gaussian distributed with mean of 0 and constant variance σ_ϵ^2.

3.1.2 Proportional Hazard Models

When interest is on an event outcome, both parametric (exponential or Weibull) and semiparametric (COX) models are available to model survival data. Let T be the time to the event, and the survival function is defined as $S(t) = P(T > t)$. Hazard rate function is a way to model the instantaneous rate of occurrence of the event in survival analysis. The chance for the event occurrence, given the individual surviving until time t, can be expressed as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t <= T < t + \Delta t) | (T >= t)}{\Delta t} = \frac{f(t)}{S(t)} = \frac{-dlog[S(t)]}{dt},\ t > 0. \tag{3.3}$$

A large family of parametric survival models focusing on hazard rate function have been applied in survival analysis. The proportional hazards model is commonly used (Cox, 1972). The hazard at time t is described using the regression model for a subject with covariates $w_i^\top$ as

$$h_i(t|w_i) = h_0(t) exp(w_i^\top \alpha), \tag{3.4}$$

in which $h_0(t)$ denotes a baseline hazard function. While no assumption about the form of $h_0(t)$ is assumed in the semiparametric Cox model, parametric models such as exponential and Weibull models are also popular. $w_i^\top \alpha$ in the exponent is the linear combination of the predictors. In this way, the predictors have a multiplicative or proportional effect on the predicted hazard $h_i(t|w_i)$ and the effect of increase or reduction on risk is the constant at all duration. The corresponding survival function is given as

$$S_i(t|w_i) = S_0(t)^{exp(w_i^\top \alpha)},\ t > 0, \tag{3.5}$$

where $S_0(t)$ is a baseline survival function. Some of the covariate values are fixed during follow-up, such as age, gender, education level. Time-varying covariate that changes over

time during the follow-up period can also be analyzed with the regression model to estimate its effect on survival time.

3.1.3 Latent Class Model

We have discussed the problem of using separate longitudinal and survival models when the two outcomes are highly correlated in the introduction chapter. The joint model approach makes it possible to properly process the information contained in both data. One approach is joint latent class model (JLCM). The classical theory of latent class analysis was first created as a method to explain the heterogeneity in response patterns in survey design with binary indicator variables (Lazarsfeld, 1950). As more and more applications are discovered, this method is no longer only considered for categorical variables, but rather as a general modeling tool to account for heterogeneity in data. The latent class model assumes that there are multiple (i.e. more than one) heterogeneous groups within a population. The heterogeneous groups so called latent classes that are believed to exist, but not directly observed, and the probability of each individual belonging to each of the latent classes can be modelled. The joint latent class model consists of three regression submodels. Following the notations developed in the above section, let Y_{ij} denote the longitudinal measurements for subject i at time t_{ij}. These n subjects are divided into G latent homogeneous subgroups. G is a finite number and usually falls within range of 2 to 6. If G=1, the whole sample is a homogeneous case. c_i is used as the class membership for subject i and $c_i = g$ if this subject is in class $g(g = 1, 2, ..., G)$.

3.1.3.1 Latent Class Probability

The class membership probability π can be described with a multinomial logistic regression model:

$$\pi_{ig} = P(c_i = g) = \frac{exp(x_i^\top \xi_g)}{\sum_{l=1}^{G} exp(x_i^\top \xi_l)}, \tag{3.6}$$

where x_i including an intercept is a vector of covariates associated with the parameter vector ξ_g. The elements in ξ_g are the coefficients of the explanatory variables for belonging to class g, thus are class specific. Since a reference class is required to ensure identifiability as $\sum_{l=1}^{G} = 1$, we choose class G as the reference with $\xi_G = 0$. Note that in practice, x_i can be reduced to only have intercept when no predictor of class membership is assumed and the regression model reduces to a class-specific probability.

Each latent class is characterized by a class-specific distribution of longitudinal marker and time to event. The key assumption is that these two outcomes are conditionally independent given these latent classes and covariates.

3.1.3.2 Latent Process for the Mixed-Effects Model

Subjects can be categorized in G latent classes with different trajectories. Given the latent class g, the conditional longitudinal outcome Y_{ij} can be expressed in a standard mixed-effects model:

$$Y_{ij}|_{c_i=g} = x_{ij}^\top \beta_g + z_{ij}^\top b_{ig} + \epsilon_{ij}, \tag{3.7}$$

where x_i is a vector of covariates for fixed effects with β_g as the corresponded regression coefficients, in which the time effect is always class specific. The fixed effects of other covariates may or may not be class specific. b_{ig} is a vector of random-effects and z_i is the

design matrix, e.g. intercept and slope. Subject level b_i captures the individual variability in the trajectories and it can have a class-specific distribution $b_{ig} = b_i|_{c_i=g} \sim N(0, \Sigma_{bg})$, in which Σ_{bg} is an unspecified variance-covariance matrix. β_g and b_{ig} together describe the shape and change with time of the longitudinal outcome in a particular g class. The measurement error term ϵ_i is following $N(0, \sigma_\epsilon^2 I)$ as homoscedastic independent errors.

3.1.3.3 Latent Process for the Risk of Event

Given the latent class g, using proportional hazard model, the risk of event can be described as:

$$h_i(t)|_{c_i=g} = h_{0g}(t)exp(w_i^\top \alpha_g), \tag{3.8}$$

where w_i is a vector of covariates associated with parameters α_g in class g. The baseline risk function h_{0_g} is specific to each latent class and can be modeled using parametric risk functions, such as Weibull or cubic M-splines, to maintain likelihood framework.

3.1.3.4 Joint Model with More Than One Longitudinal Outcome

With regard to the longitudinal submodel, one extension is to allow for more than one longitudinal outcome (i.e. multiple biomarkers), commonly known as a multivariate joint model. Based on the definition of latent class modeling, these outcomes trajectories are all class specific. This is motivated by the highly correlated clinical responses sharing the similar pattern of change within homogeneous subgroup. For example, recent studies have shown that cognitive and motor processes are functionally related, and a similar evolution across time is suggested for both (Buchman and Bennett, 2011; Kueper et al., 2017). Conditional on the latent classes, these longitudinal processes are independent. Here, we go beyond the

standard joint model that deals with single longitudinal outcome to a model with bivariate outcomes. We consider two longitudinal markers indicated with k=1 and 2, each evolution is described by a mixed model specific to each class g. Assume Y_{ijk} corresponds to the measure of k^{th} longitudinal marker observed at the j^{th} time point for i^{th} subject, we have class specific longitudinal submodel:

$$Y_{ijk}|_{c_i=g} = x_{ijk}^{\top}\beta_{gk} + z_{ijk}^{\top}b_{igk} + \epsilon_{ijgk},\ (k = 1, 2), \tag{3.9}$$

where x_{ijk} and z_{ijk} are vectors of covariates associated with class specific fixed effect parameters β_{gk} and class specific random effects parameter b_{igk}.

One previous study with latent class approach accommodated multiple longitudinal biomarkers in the joint model by using a multivariate longitudinal model in which same factors predicting all the outcomes (Jacqmin-Gadda et al., 2009) and the analysis more focused on the overall evolution of longitudinal profile rather than a specific marker. In contrast, in our proposed model, each longitudinal marker is considered as a independent outcome conditional on the latent class, and different covariates or predictors (x_i and z_i) are allowed to differ over the k longitudinal submodels. Specific choice of the link function based on different distributed outcome can also be applied, for example, a logistic or Poisson regression. The estimation of such computation intensive model can be achieved by using Bayesian approach which is has been barely used in the joint latent class model and will be discussed in next section.

3.1.4 Estimation and Assessment

Different estimation approaches have been utilized to fit the latent class models. For the JLCM, the published works mostly relied on maximum-likelihood estimation (MLE).

The difficulties when using MLE are the problem of multiple local maxima and expensive computation burden as the number of data dimensions increases. Bayesian method could be provided as an alternative for inferences in JLCM. Compared to MLE, Bayesian joint model approach allows a more straightforward model assessment without the potential numerical integration in high dimensions.

Under the conditional independence assumption, inside each latent class, the risk of event is independent of the measured mark evolution with adjustment of covariates. The likelihood for each subject can be expressed as:

$$L_i(\theta_G) = \sum_{g=1}^{G} f(Y_{ik}|c_i = g; \theta_G) f(T_i|c_i = g; \theta_G) P(c_i = g|\theta_G), \tag{3.10}$$

where $f(Y_k|c_{i=g}; \theta_G)$ is the density of the longitudinal model for the k^{th} outcome in class g with mean $x_{ik}^\top \beta_{gk} + z_{ik}^\top b_{igk}$ and covariance matrix $z_{ik}\Sigma_{bk}z_{ik}^\top + \sigma_\epsilon^2 I$; $f(T_i|c_{i=g}; \theta_G)$ is the corresponding class-specific hazard regression function; and $P(c_i = g|\theta_G) = \pi_{ig}$ is probability being in the g^{th} class. The overall log-likelihood is $l(\theta_G) = \sum_{i=1}^{N} log(L_i(\theta_G))$.

In the Bayesian approach, parameters are treated as random variables and the distribution of a parameter is called prior. Given a fixed number of classes, the parameter vector θ_G contains elements $(\xi_g^\top, \beta_g^\top, \Sigma_{bg}, \alpha_g^\top, \sigma_\epsilon)_{g=1-G}$. A joint prior $p(\theta)$ can be specified by taking the product of priors, which are assumed to be independent of each other. Specification of prior distribution plays an important role in Bayesian inference. When priors are non-informative, then Bayesian estimates would be comparable to those obtained from a non-Bayesian algorithm such as a MLE algorithm.

In order to make the estimates driven by the data, priors are signed with vague distribution and have minor impact on posterior inferences. Usually, normal distribution prior is considered for each element the main effects vector $\beta_g^\top$ with $N(0, 1000)$ and $\alpha_g^\top$ with $N(0, 1000)$;

inverse gamma distribution is for error term variance $\sigma_\epsilon^2 \sim gamma^{-1}(0.001, 0.001)$. For random effects variance-covariance matrix Σ_{bg}, inverse-Wishart (inverse multivariate gamma) distribution with 2 degree of freedom is commonly used $b_{ig} \sim Wishart^{-1}(2, R)$, where $R = diag(0.001, 2)$. Subjects are modeled as belonging to the latent class for which they have the highest class membership probabilities. The proportion of individuals in class g can be represented by a multinomial logistic regression (formula 3.6) with class-specific coefficient ξ_g and covariate x_i on group membership. Normal distributed priors are assigned for $\xi_{1-(G-1)} \sim N(0, 1000)$. For identiability, $\xi_G = 0$. Another common prior distribution used for the latent class proportions in Bayesian is the Dirichlet distribution (Asparouhov and Muthen, 2011): $\pi_{1-G} \sim Dirichlet(\delta_1, \delta_2, ..., \delta_G)$, with density $\frac{\Gamma(\delta_1+...+\delta_G)}{\Gamma(\delta_1)...\Gamma(\delta_G)}\pi^{(\delta_1-1)}...\pi^{(\delta_G-1)}$, where proportions (π_g) sum to 1 and δ_g element represents the hyper-parameters regarding the size of the g latent class. The values of δ_{1-G} can be determined based on the knowledge on the class membership distribution from previous studies. However, most of the time, information about the class membership distribution is not known and weakly informative prior is assigned with the δs all equal to a small positive number. Nasserinejad et al., 2017 suggested that δs equal to a number slightly less than half of s, where s is the number of class-specific parameters. This setting indicates that each class must technically exist but little additional information about final class size is provided through the prior (Nasserinejad et al., 2017; Andrinopoulo et al., 2020).

Base on the probability density and priors, the joint posterior probability distribution for $p(\theta|Y, T)$ with a Dirichlet prior is

$$\begin{aligned} p(\theta|Y,T) &\propto p(Y,T|\theta)p(\theta)p(\pi) \\ &= p(Y|b_g, \beta_g, \sigma_{\epsilon g}, \pi_g)p(T|\alpha_g, \pi_g)p(b_g|\sigma_{\epsilon g})p(\sigma_{\epsilon g})p(\beta_g)p(\alpha_g)p(g|\pi_g)p(\pi_g). \end{aligned} \tag{3.11}$$

Using a multinomial logistic regression model for the membership probability, the joint posterior density can be written like

$$\begin{aligned} p(\theta|Y,T) &\propto p(Y|\theta)p(T|\theta)p(\theta)p(g|\pi)p(\xi) \\ &= p(Y|b_g,\beta_g,\sigma_{\epsilon g},\pi_g)p(T|\alpha_g,\pi_g)p(b_g|\sigma_{\epsilon g})p(\sigma_{\epsilon g})p(\beta_g)p(\alpha_p)p(g|\pi_g)p(\xi_g). \end{aligned} \tag{3.12}$$

Here, $p(Y|b_g,\beta_g,\sigma_{\epsilon g},\pi_g)$ can be further expanded to $p(Y_1|b_{1g},\beta_{1g},\sigma_{\epsilon 1g},\pi_{1g})p(Y_2|b_{2g},\beta_{2g},\sigma_{\epsilon 2g},\pi_{2g})$ for bivariate longitudinal outcomes.

The Markov chain Monte-Carlo (MCMC) method can be implemented based on the Gibbs sampler which obtains draws from the posterior distribution. In the process, posterior distribution for π_i follows either multinomial regression or Dirichlet distribution with $D(\delta_1 + n_1, ..., \delta_G + n_G)$ and is estimated over all observations; sampling of $[Y_i, T_i|\theta_g]$ is performed separately for each class. Label switching is a well-known issue in mixture latent class models. There are two main approaches to deal with this problem. One approach is to impose an artificial identifiable constraint on parameters (Celeux et al., 2000); another approach is to employ re-labelling algorithm in each iteration to minimize the posterior expectation of some loss function of the model parameters (Stephens, 2000). The label-switching problem is not observed using the multinormial regression for the membership distribution.

The initial values of the main effect parameters $\beta_g s$, $\alpha_g s$ and random effects variance matrix Σ_{bg} can be set properly in order to speed up convergence. Convergence of the sampler is assessed informally via visual checks of trace and density plots. Posterior means and standard deviations of the parameters are computed based on these samples after discarding burn-in and summarized. As the choice of the number of latent classes is a challenging question, decision on the final model should consider the model goodness of fit, reasonable latent class sizes and meaningful interpretation of the differences between classes. The prediction of the subject's latent class is based on posterior class membership probability. Some advantages of using Bayesian analysis is its ability to incorporate prior information if

some subjects' latent classes are known or important classification information is available from external sources.

In practice, models are repeatedly estimated with different numbers of classes. We assume an overfitted model with superfluous latent classes will converged to the true model by eliminating the unnecessary classes and this is supported by the study of Nasserinejad et al., 2017 and Andrinopoulo et al., 2020. The best model is selected with information criteria. Widely Applicable Information Criterion (WAIC), also known as the Watanable-Akaike for Bayesian model selection (Watanabe, 2010) is particularly helpful for models with hierarchical and mixture structures. This criterion is composed of the expected mean value over the posterior distribution and the effective number of parameters:

$$WAIC = -1/n \sum_{i=1}^{n} logE[p(\theta)] + 1/n \sum_{i=1}^{n} \{E_\theta[logp(\theta)]^2 - E_\theta^2[logp(\theta)]\}. \tag{3.13}$$

It sums each data point value based on the point-wise posterior expected density distribution, and it is considered to be more stable compared to DIC. Besides WAIC, relative entropy is considered as an indicator for a good separation of the identified classes. Relative entropy is calculated as $1 + \frac{\sum_i^N \sum_g^G \hat{\pi}_{ig} log(\hat{\pi}_{ig})}{NlogG}$, where $\hat{\pi}_{ig}$ is the estimated posterior probability of individual being in the latent class g. As the value of entropy decreases, the classes become less distinguished from one to another. Entropy is used to rule out the number of classes with a cut-off point, for example, below 0.6 (The cut-off value depends on the data and the model complexity.) (Asparouhov and Muthen, 2014). The more discriminatory the posterior classification is, the better the model.

So far there are not many software packages provided for fitting JCLM. Very recently, a package "lcmm" (Proust-Lima et al., 2017) in R software has been developed and JLCM can be fitted through the $Jointlcmm()$ function in this package. The estimation is based on MLE using Marquardt algorithm (MA). Due to the occurrence of multiple local maxima in

MLE, iterative estimation MA needs to be initialized with different initial values for θ_G to ensure convergence towards the global maximum. This package does not apply on the joint model with multivariate longitudinal outcomes.

3.2 Simulation

We conduct a simulation study evaluating the performance of the Bayesian joint latent class modelling approach. We consider a model for bivaraite longitudinal markers, one event and two latent classes. Parameter values are chosen to mimic the application and the data are generated under the following settings:

The longitudinal submodel:

$Y_{ik}|_{ci=g} = \beta_{0kg} + b_{i0} + \beta_{1kg} \times t + \epsilon_{ik}$, and $k = 1$ *or* 2

with $\beta_{011} = 1.0, \beta_{111} = -0.02, \beta_{021} = 2.0$, and $\beta_{121} = -0.05$ in class 1;

$\beta_{012} = 2.0, \beta_{112} = -0.1, \beta_{022} = 4.0$, and $\beta_{122} = -0.2$ in class 2;

$b_{i01} \sim N(0, 0.1), b_{i02} \sim N(0, 0.2), \epsilon_{i1} \sim N(0, 0.05)$ and $\epsilon_{i2} \sim N(0, 0.2)$ in both classes.

The survival submodel:

$h(t)|_{ci=g} = h_0(t)exp(\alpha_{0g} + \alpha_{1g}x)$

with $\alpha_{01} = -4.2, \alpha_{11} = 0.2$ in class 1;

$\alpha_{02} = -2$, and $\alpha_{12} = 0.1$ in class 2;

Covariate $x \sim Bernoulli(0.5)$.

We have 500 simulated datasets and each contains 400 subjects. For each subject, the membership is determined according to a Bernoulli distribution with probability $\pi_1 = 0.70$ for class 1 and $\pi_2 = 30$ for class 2. The visit time is set as 0 to 6 (total 7 visits with equal interval). The measurements of Y_1 and Y_2 share the same frequency and collecting time for the same subject. The time of event is determined by the Weibull function in the latent

class assuming all the subjects eventually will experience the event. Subjects with event happening after the last visit is considered as censored and longitudinal measures after the onset of the event are excluded. Based on the setting, we have 92.0% censored subjects with mean follow-up time 5.7 at censoring or event in the first class and 40.3% censored with follow-up time 4.7 in the second class.

We use both multinomial regression without covariate and Dirichlet process prior to model the class membership probability in the proposed joint model. We fit our Bayesian model for each simulated dataset by running a single chain MCMC using WinBUGS and the "R2WinBUGS" package (Sturtz et al., 2005) to invoke WinBUGS from R for Bayesian analysis. A type of label switching that class switches over replications is considered during the simulation. Relabeling algorithm is applied by comparing the parameter estimates with generating values given the true values of the parameters are known in each class and correction can be made (Cho et al, 2013). The results summarized in Table I show that the estimator is fairly good in terms of posterior mean, standard deviation and percentage of coverage using either multinomial regression method or Dirichlet process for class membership probability.

3.3 Application to Memory and Aging Project Cohort

We apply the latent class joint modeling method to the Rush MAP cohort, a longitudinal study of common chronic conditions in old age. Subjects selected in this analysis are without any type of dementia or mild cognitive impairment at baseline. A small number of subjects are excluded due to lack of basic demographic information, for example, age, gender, educational level etc. We further exclude people who have less than four years of follow-up visits. Longitudinal measurements after the time point when a person was diagnosed with AD are not used. The final sample includes 717 participants with a mean age of 78.2 $\pm$ 6.9 years at baseline, 77.1% female, 15.0 $\pm$ 3.0 years of education, BMI 27.4 $\pm$ 5.2 at baseline

Table I: SIMULATION STUDY RESULT FOR JOINT LATENT CLASS MODEL OF BIVARIATE LONGITUDINAL OUTCOMES AND AN EVENT

		Multinomial regression[a]			Dirichlet prior[b]		
Parameter	True value	Posterior mean	SD	CR%	Posterior mean	SD	CR%
Class membership							
π_1	0.70	0.70	NA	NA	0.70	NA	NA
Parameters in class 1							
β_{011}	1.0	1.000	0.006	93.0	1.000	0.007	92.8
β_{111}	-0.02	-0.020	0.001	95.4	-0.020	0.001	95.0
β_{021}	2.0	2.000	0.030	93.4	1.999	0.030	93.2
β_{121}	-0.05	-0.050	0.006	94.8	-0.050	0.006	94.0
α_{01}	-4.2	-4.311	0.395	91.4	-4.317	0.384	91.2
α_{11}	0.2	0.219	0.452	92.0	0.222	0.450	92.4
Parameters in class 2							
β_{012}	2.0	2.000	0.012	90.8	2.000	0.012	91.4
β_{112}	-0.1	-0.100	0.001	95.2	-0.100	0.001	95.4
β_{022}	4.0	4.000	0.050	94.6	4.002	0.051	95.4
β_{122}	-0.2	-0.200	0.012	95.4	-0.200	0.012	94.8
α_{02}	-2	2.045	0.241	94.0	-2.046	0.238	92.6
α_{12}	0.1	1.107	0.246	95.6	0.106	0.246	96.0
Common parameters							
$\sigma^2(b_{i01})$	0.1	0.100	0.007	93.8	0.100	0.006	95.0
$\sigma^2(b_{i02})$	0.2	0.198	0.020	92.4	0.199	0.020	93.6
$\sigma^2(\epsilon_{i1})$	0.05	0.050	0.004	94.2	0.050	0.003	94.0
$\sigma^2(\epsilon_{i2})$	0.5	0.501	0.009	93.8	0.501	0.009	93.4

[a] Multinomial logistic regression modelling the class membership probability.
[b] Dirichlet prior for the class membership probability.

and 20.8% ApoE4 carriers. The average follow-up time is 9.7 ± 3.5 year with a total of 132 subjects (18.4%) diagnosed with AD during the follow-up. Our objective is to jointly model the trajectories of cognition related outcomes with AD risk conditional on the latent class.

3.3.1 Mixed-Effects Model for the Trajectories of the Cognitive Function and Related Outcomes

Our main outcome in the longitudinal analysis is *Globcog*, an annual measure of global cognitive performance derived by averaging the z-scores from 19 annual tests. Another outcome of interest is the motor function $Motor$, a composite measure constructed by converting the raw scores from 10 motor measures. There are many reports of non-cognitive symptoms, e.g. loss of motor function, accompanied with cognitive decline during the long AD preclinical phase. Also, body mass index (BMI) of all participants were collected at each evaluation. It has been reported that weight loss precedes dementia diagnosis and weight loss is an early marker for AD dementia disorder (Yang and Gao, 2013). Changes in BMI are associated with increased risk of dementia (Alhurani et al., 2016).

We first like to show the trend of these longitudinal measures prior to clinical AD diagnosis using separate mixed effects model. Since these outcomes were collected at the same visit in MAP cohort, the time variable is shared in the data. Time is recorded as 0, 1, ..., in years until the last follow-up or the time when AD was diagnosed. For covariates, *age_bl* (in year) is at the date of the first cognitive assessment (baseline); *gender* is coded as 0 for female and 1 for male; *edu* (in year) is based on self-reported years of regular schooling; and *ApoE*4 (genotyping, 1 for positive and 0 for negative) is based a high throughput sequencing. BMI_bl is BMI at baseline. These covariates are well known factors to the cognition or motor impairment and commonly included in the analysis of AD data. BMI as an outcome or covariate is centered by subtracting 27.

Classical longitudinal mixed effects model with random intercept and slope is applied for the linear trend of these outcomes as:

$$Globcog_{ij} = \beta_0 + \beta_1 \times time_{ij} + \beta_2 \times age\_bl_i + \beta_3 \times gender_i + \beta_4 \times edu_i + \beta_5 \times ApoE4_i$$
$$+ b_{0i} + b_{1i} \times time_{ij} + \epsilon_{ij}$$
$$Motor_{ij} = \beta_0 + \beta_1 \times time_{ij} + \beta_2 \times age\_bl_i + \beta_3 \times gender_i + \beta_4 \times edu_i + \beta_5 \times BMI\_bl_i$$
$$+ b_{0i} + b_{1i} \times time_{ij} + \epsilon_{ij}$$
$$BMI_{ij} = \beta_0 + \beta_1 \times time_{ij} + \beta_2 \times age\_bl_i + \beta_3 \times gender_i$$
$$+ b_{0i} + b_{1i} \times time_{ij} + \epsilon_{ij}$$

The results are summarized in the top panel of Table II The estimated average regression coefficient of time is -0.047 with 95% confidence interval of (-0.052, -0.042) for cognitive decline, the coefficient of time is -0.029 with confidence interval of (-0.031, -0.028) for motor function score and -0.098 with confidence interval of (-0.124, -0.072) for BMI, suggesting these three measures all decline significantly over the study period. The random intercept and slope variances estimates in these models shown as diagonal elements in Σ_b 2×2 covariance matrix are all significant with confidence interval not covering 0.

3.3.2 Proportional Hazards Model for AD risk

In this part, we start with a simple proportional hazards survival function with Weibull baseline hazard. Subjects diagnosed with AD at baseline are excluded, resulting in a sample size of 717, of which 132 (18.4%) developed AD later. With linear function of covariates (age, gender, education and ApoE4) at baseline level, hazard h at time t for individual i is:

$$h_i(t) = h_{0i}(t) exp(\alpha_0 + \alpha_1 \times age\_bl_i + \alpha_2 \times gender_i + \alpha_3 \times edu_i + \alpha_4 \times ApoE4_i) \ .$$

Table II: PARAMETER ESTIMATES FROM SEPARATE MODELS ON COGNITION, MOTOR FUNCTION, BMI DECLINE AND AD RISK

Parameter	Estimate	95% CI
Longtitudial mixed-effects model		
Globcog		
Intercept (β_0)	0.941	(0.608, 1.273)
Time (β_1)	-0.047	(-0.052, -0.042)
Age_bl (β_2)	-0.017	(-0.021, -0.013)
Gender (β_3)	-0.071	(-0.133, -0.009)
Edu (β_4)	0.054	(0.046, 0.063)
ApoE4 (β_5)	0.043	(-0.021, 0.106)
Ran_int (σ_0^2)	0.107	(0.093, 0.120)
Ran_time (σ_1^2)	0.003	(0.003, 0.004)
Ran_cov (Cov_{01})	-0.000	(-0.002, 0.002)
Error (σ_ϵ^2)	0.048	(0.048, 0.049)
Motor		
Intercept (β_0)	2.802	(2.630, 2.975)
Time (β_1)	-0.029	(-0.031, -0.028)
Age_bl (β_2)	-0.019	(-0.021, -0.018)
Gender (β_3)	0.063	(0.036, 0.090)
Edu (β_4)	0.006	(0.002, 0.009)
BMI (β_5)	-0.010	(-0.012, -0.008)
Ran_int (σ_0^2)	0.023	(0.020, 0.027)
Ran_time (σ_1^2)	0.001	(0.001, 0.002)
Ran_cov (Cov_{01})	-0.001	(-0.001, -0.000)
Error (σ_ϵ^2)	0.013	(0.012, 0.014)
BMI		
Intercept (β_0)	14.069	(9.971, 18.167)
Time (β_1)	-0.098	(-0.124, -0.072)
Age_bl (β_2)	-0.170	(-0.222, -0.118)
Gender (β_3)	-0.291	(-1.148, 0.565)
Ran_int (σ_0^2)	25.247	(22.897, 27.596)
Ran_time (σ_1^2)	0.079	(0.067, 0.090)
Ran_cov (Cov_{01})	-0.347	(-0.469, -0.225)
Error (σ_ϵ^2)	2.951	(2.855, 3.046)
Proportional hazards model		
Intercept (α_0)	-2.880	(-2.960, -2.877)
Age_bl (α_1)	0.043	(0.034, 0.052)
Gender (α_2)	-0.028	(-0.154, 0.088)
Edu (α_3)	-0.013	(-0.029, 0.003)
ApoE4 (α_4)	0.174	(0.066, 0.280)

Assumption of proportional hazards is tested using a score test based on scaled Schoenfeld residuals (Grambsch and Therneau, 1994). The results are summarized in Table II (bottom). Our data shows that age at baseline ($\beta = 0.043$), ApoE4 allele ($\beta = 0.174$) are strong risk factors for AD. Gender has no effect. Education level is found to be negatively associated ($\beta = -0.013$) with AD incidence but the effect is not significant. Earlier studies have reported negative association between AD onset and having a higher level of education. Interestingly, more recent research is showing that higher education level does not prevent people from a delay in developing AD (Wilson et al., 2019).

3.3.3 Joint Latent Class Modeling of Cognition Related Outcome Decline and AD risk

In the MAP cohort, the data suggest that age-related cognitive and motor decline are related. To the best of our knowledge, there has been no analysis jointly modelling both cognition and motor functions ($Globcog\&Motor$) longitudinally with risk of AD. In the same time, we also like to jointly model cognition and BMI ($Globcog\&BMI$) changes across time. Here we aim to identify subgroups of MAP participants who are similar with regard to their longitudinal changes in aging related function markers and measures. Moreover, we like to know if these markers' trajectories are related to their risk of AD in a clinical meaningful way.

In the implementation of the JLCM, for longitudinal outcomes, we are mostly interested in the linear trend over time and the trajectories are set to be specific to each latent class while other covariates effects, variances of random effects and measurement errors are common for all classes. In the proportion hazard submodel, the intercept and effect of age at baseline are considered to be class specific. The baseline risk function, the effects of gender, education and ApoE4 are common across classes. No covariates are included in the class membership model.

With this setting, we have parameters $\beta_g = (\beta_{int(g)}, \beta_{time(g)}, \beta_{age\_bl}, \beta_{gender}, \beta_{edu}, \beta_{ApoE4}, \beta_{BMI})$ for fixed effects and $\Sigma_b, \sigma_\epsilon$ for variances covariances in longitudinal submodel; γ_{Weib}, $\alpha_g = (\alpha_{int(g)}, \alpha_{age\_bl(g)}, \alpha_{gender}, \alpha_{edu}, \alpha_{ApoE4})$ for risk effects in proportional hazard submodel; $\xi_{int(g)}$ in the multinomial logistic regression model for class-membership probability. We run MCMC in WinBUGS with vague priors to obtain the parameter estimates. Trace and density plots are presented in the Appendix B.

We first compare models with varying number of latent classes. We start with model I without classification. Corresponding to our setting for the class-specific parameters: one intercept in the class-membership model ($\xi_{int(g)}$), two fixed effects in the linear trajectory of intercept and time ($\beta_{int(g)}$ and $\beta_{time(g)}$) for each of the longitudinal outcome, two fixed effects in hazard model ($\alpha_{int(g)}$ and $\alpha_{age\_fb(g)}$), together 7 parameters are introduced to the model when an additional latent class is considered. We only show the results for model with up to 3 latent classes because one class tends to be empty (<3%) when total 4 classes are assigned. This will be discussed in next the Discussion part.

Table III shows Model I to III with number of latent classes, number of parameters, deviance, WAIC, entropy and the posterior proportion of each class for models of $Globcog\&Motor$ and $Globcog\&BMI$. As class number increases, the fitting shows improvement with decreased WAIC. Comparing to Model I and II, Model III containing 3 classes yields the smallest WAIC for both sets of outcomes. To further investigate whether model III gives the best fit to our data, we check the posterior class membership probability to assess the model goodness of fit and class discrimination. The posterior class membership probability for subject i in latent class g is $\pi_{ig} = P(c_i = g|Y_{1i}, Y_{2i}, T_i, x_{cov_i}, \hat{\theta}_G)$. The relative entropy of Model III calculated based on the individual posterior probabilities of each class membership is 0.77 and 0.86 for model $Globcog\&Motor$ and $Globcog\&BMI$, respectively, suggesting that model with 3 classes provides good discrimination for both joint models. A posterior classification is computed by assigning subjects to the class where they have the highest posterior

probabilities in class membership ($\hat{c_i} = max(\pi_{ig})$). The estimated mean probability of an individual being in each class stratified by class membership $max(\pi_{ig})$ is summarized in Table IV. Model III for $Globcog\&Motor$ is comprised of 3 classes with 54.0%, 29.8% and 16.2% of total subjects allocated and the mean of the posterior probabilities in each class ranges with 92.0% (class 1), 77.2% (class 2) and 85.3% (class 3) shown as the diagonal numbers. Model III for $Globcog\&BMI$ with classes containing 40.0%, 21.5% and 38.5% of sample has the mean of the posterior probabilities in each class with 91.7% (class 1), 96.1% (class 2) and 88.8% (class 3). And most off-diagonal numbers are less than or close to 10% indicating discrimination between classes is correct and classification is clear. Model of cognition and BMI is slightly better than the model of cognition and motor function in term of class discrimination.

Table III: MODEL COMPARISON FOR VARYING NUMBERS OF LATENT CLASSES

Model	G	npm	deviance	WAIC	Entropy	%class1	%class2	%class3
Cognition and Motor								
Model I	1	26	-9772	-7984	NA	100		
Model II	2	33	-9859	-8060	0.83	44.5	55.5	
Model III	3	40	-10150.0	-8257	0.77	54.0	29.8	16.2
Cognition and BMI								
Model I	1	24	27940	29869	NA	100		
Model II	2	31	27880	29842	0.71	59.2	40.8	
Model III	3	38	27695	29685	0.86	40.0	21.5	38.5

Table V and VI show the posterior means and 95% credible intervals of the parameters in the longitudinal submodel for $Globcog\&Motor$ and $Globcog\&BMI$, respectively, with survival submodel for AD risk. The intercept terms (ξ_1 and ξ_2) in the class membership probability models are both significant in both models. With adjustment of covariates, the effects of time are all significantly different from 0 for these longitudinal outcomes in all latent classes. In Table V, majority of the sample (class 1, 54.0%) shows "a natural aging status" with a relatively slow cognitive decline rate ($\beta_{111} = -0.015$); while the smallest group (class

Table IV: MEAN POSTERIOR PROBABILITIES OF CLASS MEMBERSHIP

Class membership[a]	Prob/Class1	Prob/Class2	Prob/Class3
Cognition and Motor			
Class1 (54.0%)	0.920	0.078	0.002
Class2 (29.8%)	0.157	0.772	0.071
Class3 (16.2%)	0.011	0.137	0.853
Cognition and BMI			
Class1 (40.0%)	0.917	0.081	0.002
Class2 (21.5%)	0.005	0.961	0.034
Class3 (38.5%)	0.101	0.011	0.888

[a] Membership is determined by an individual's highest class-specific probability.

3) with 16.2% of the sample shows the fastest decline rate of cognition ($\beta_{113} = -0.162$). Class 2 (29.8%) is characterized as "intermediate" with a more pronounced cognitive decline ($\beta_{112} = -0.056$) than that of class 1, but not as dramatic as class 3. We have the same trend with having the slowest decline rate ($\beta_{211} = -0.027$) in Class 1 and the fastest decline ($\beta_{213} = -0.036$) in class 3 for motor function. However, the slope differences are not as dramatic as those in cognitive decline, especially for the slopes between class 1 and class 2. In AD risk submodel, age at baseline has significant positive association with AD risk in all classes. The age effect shown as regression coefficient becomes weaker ($\alpha_{11} = 0.43$ in class 1 and $\alpha_{13} = 0.20$ in class 3) as cognition and motor functions decline faster.

In model of cognition and BMI (Table VI), the estimated intercepts and slopes for the two outcomes are all different and distinguishable across classes. We see class 1 contains 40.0% of total subjects and is characterized by relative "normal aging/stable BMI" ($\beta_{111} = -0.025$ and $\beta_{211} = -0.032$) starting with the highest cognitive score (0.58) and relatively normal BMI (24.7). Class 2 with 21.5% of the sample is categorised by "intermediate declining cognition/BMI" ($\beta_{112} = -0.046$ and $\beta_{212} = -0.101$). We also notice that the starting point of BMI is the highest (34.2) among the three classes. The last class with "rapid declining cognition/BMI" ($\beta_{113} = -0.071$ and $\beta_{213} = -0.156$) covers 38.5% of the sample and cognition starts from a low value (0.22). In the survival model, age at entry has a

positive and significant effect on AD risk in all classes, and the hazard ratio per year older varies from 1.1 to 1.2 among classes.

Figure 3 provides the predicted class-specific trajectories of *Globcog&Motor* function scores (top) and *Globcog&BMI* (bottom) over years with the corresponding survival function (AD-free probability) of AD. Both models have 3 latent classes, but class members could be different between the two models. The "stable" class representing the healthiest group (Class 1, green) has the best AD-free rate. This class is also characterized by the highest starting point for cognition and motor function and the lowest for BMI value. The "intermediate" class (Class 2, blue) has a relative lower rate of AD diagnoses. The highest rate of AD incidence is observed in the "rapid decline" class (Class 3, red). In the model of *Globcog&Motor*, the incidence of AD in this class starts right at 5 years of follow up and survival function reaches the bottom after 10 years, meaning that at this time all the subjects in this class are roughly diagnosed as AD. In the model of *Globcog&BMI*, the survival curves in the three classes are different but not as dramatic as in the model of *Globcog&Motor*.

When we compare the subjects' base characteristics among the 3 latent classes, we find the classes are significantly different according to age at entry in both models (Appendix C). Class 1, the healthiest group, includes mainly younger subjects while class 3 with the fastest decline rate has relatively older subjects. No significant difference is observed from gender or education level. In the model of *Globcog&Motor*, compared with the class 1, class 2 and class 3 (fastest decline)'s subjects have relatively higher percentage of ApoE4 carriers and the difference is significant.

Above analysis setting is using multinomial logistic regression without covariate predicting the latent class membership. We repeat the analysis using weakly informative Dirichlet prior for the class proportion with equal weights $D(\delta_1, ..., \delta_G)$. In particular, following the recommendation in Nasserinejad et al., 2017, we set all δ equal to 3. As expected, the re-

Table V: PARAMETER ESTIMATES OF MODEL III FOR COGNITION AND MOTOR FUNCTION (3 CLASSES)

Parameter	Parameter label	Posterior mean	95% CI
Membership			
int class 1	ξ_1	-0.560	(-0.745, -0.506)
int class 2	ξ_2	-1.225	(-1.432, -1.116)
Cognition			
Intercept class1	β_{101}	0.468	(0.421, 0.514)
Intercept class2	β_{102}	0.332	(0.273, 0.398)
Intercept class3	β_{103}	0.383	(0.301, 0.469)
Time class1	β_{111}	-0.015	(-0.020, -0.011)
Time class2	β_{112}	-0.056	(-0.065, -0.049)
Time class3	β_{113}	-0.162	(-0.175, -0.149)
Age_bl	β_{12}	-0.018	(-0.022, -0.014)
Gender	β_{13}	-0.067	(-0.128, -0.005)
Edu	β_{14}	0.055	(0.046, 0.063)
ApoE4	β_{15}	0.031	(-0.034, 0.097)
Var(Intercept)	σ^2_{10}	0.104	(0.093, 0.118)
Var(Time)	σ^2_{11}	0.001	(0.001, 0.001)
Cov(Intercept_time)	Cov_{101}	-0.002	(-0.004, -0.001)
Error	$\sigma^2_{\epsilon 1}$	0.220	(0.216, 0.224)
Motor			
Intercept class1	β_{201}	1.362	(1.308, 1.420)
Intercept class2	β_{202}	1.297	(1,236, 1.361)
Intercept class3	β_{203}	1.282	(1.220, 1.349)
Time class1	β_{211}	-0.027	(-0.029, -0.025)
Time class2	β_{212}	-0.030	(-0.033, -0.027)
Time class3	β_{213}	-0.036	(-0.040, -0.031)
Age_bl	β_{22}	-0.018	(-0.020, -0.016)
Gender	β_{23}	0.063	(0.036, 0.090)
Edu	β_{24}	0.005	(0.002, 0.009)
BMI	β_{25}	-0.010	(-0.012, -0.008)
Var(Intercept)	σ^2_{20}	0.023	(0.020, 0.026)
Var(Time)	σ^2_{21}	0.000	(0.000, 0.000)
Cov(Intercept_time)	Cov_{201}	-0.001	(-0.001, -0.000)
Error	$\sigma^2_{\epsilon 2}$	0.115	(0.113, 0.117)
AD risk			
Intercept class1	α_{01}	-3.088	(-3.222, -2.978)
Intercept class2	α_{02}	-2.729	(-2.797, -2.688)
Intercept class3	α_{03}	-2.302	(-2.367, -2.238)
Age class1	α_{11}	0.043	(0.022, 0.058)
Age class2	α_{12}	0.037	(0.029, 0.045)
Age class3	α_{13}	0.020	(0.011, 0.030)
Gender	α_2	0.300	(-0.141, 0.745)
Edu	α_3	-0.006	(-0.016, 0.005)
ApoE4	α_4	0.001	(-0.080, 0.080)

Table VI: PARAMETER ESTIMATES OF MODEL III FOR COGNITION FUNCTION AND BMI (3 CLASSES)

Parameter	Parameter label	Posterior mean	95% CI
Membership			
int class 1	ξ_1	-0.560	(-0.745, -0.506)
int class 2	ξ_2	-1.183	(-1.445, -0.937)
Cognition			
Intercept class1	β_{101}	0.580	(0.533, 0.628)
Intercept class2	β_{102}	0.434	(0.383, 0.484)
Intercept class3	β_{103}	0.221	(0.171, 0.269)
Time class1	β_{111}	-0.025	(-0.031, -0.016)
Time class2	β_{112}	-0.046	(-0.055, -0.036)
Time class3	β_{113}	-0.071	(-0.080, -0.061)
Age_bl	β_{12}	-0.018	(-0.022, -0.014)
Gender	β_{13}	-0.078	(-0.123, -0.032)
Edu	β_{14}	0.053	(0.047, 0.058)
ApoE4	β_{15}	0.002	(-0.057, 0.061)
Var(Intercept)	σ^2_{10}	0.083	(0.072, 0.097)
Var(Time)	σ^2_{11}	0.003	(0.003, 0.004)
Cov(Intercept_time)	Cov_{101}	-0.003	(-0.005, -0.002)
Error	$\sigma^2_{\epsilon 1}$	0.220	(0.216, 0.224)
BMI			
Intercept class1	β_{201}	-2.257	(-2.679, -1.835)
Intercept class2	β_{202}	7.195	(6.704, 7.691)
Intercept class3	β_{203}	-1.183	(-1.617, -0.750)
Time class1	β_{211}	-0.032	(-0.073, -0.008)
Time class2	β_{212}	-0.101	(-0.15, -0.051)
Time class3	β_{213}	-0.156	(-0.196, -0.107)
Age_bl	β_{22}	-0.137	(-0.173, -0.106)
Edu	β_{23}	-0.182	(-0.255, -0.114)
Var(Intercept)	σ^2_{20}	9.686	(8.616, 10.860)
Var(Time)	σ^2_{21}	0.065	(-0.054, -0.077)
Cov(Intercept_time)	Cov_{201}	-0.234	(-0.321, -0.154)
Error	$\sigma^2_{\epsilon 2}$	1.732	(1.702, 1.764)
AD risk			
Intercept class1	α_{01}	-3.193	(-3.361, -3.050)
Intercept class2	α_{02}	-2.885	(-2.998, -2.784)
Intercept class3	α_{03}	-2.711	(-2.792, -2.640)
Age class1	α_{11}	0.045	(0.025, 0.067)
Age class2	α_{12}	0.034	(0.023, 0.044)
Age class3	α_{12}	0.035	(0.017, 0.054)
Gender	α_2	0.220	(-0.340, 0.767)
Edu	α_3	-0.006	(-0.022, 0.008)
ApoE4	α_4	0.252	(0.0152, 0.352)

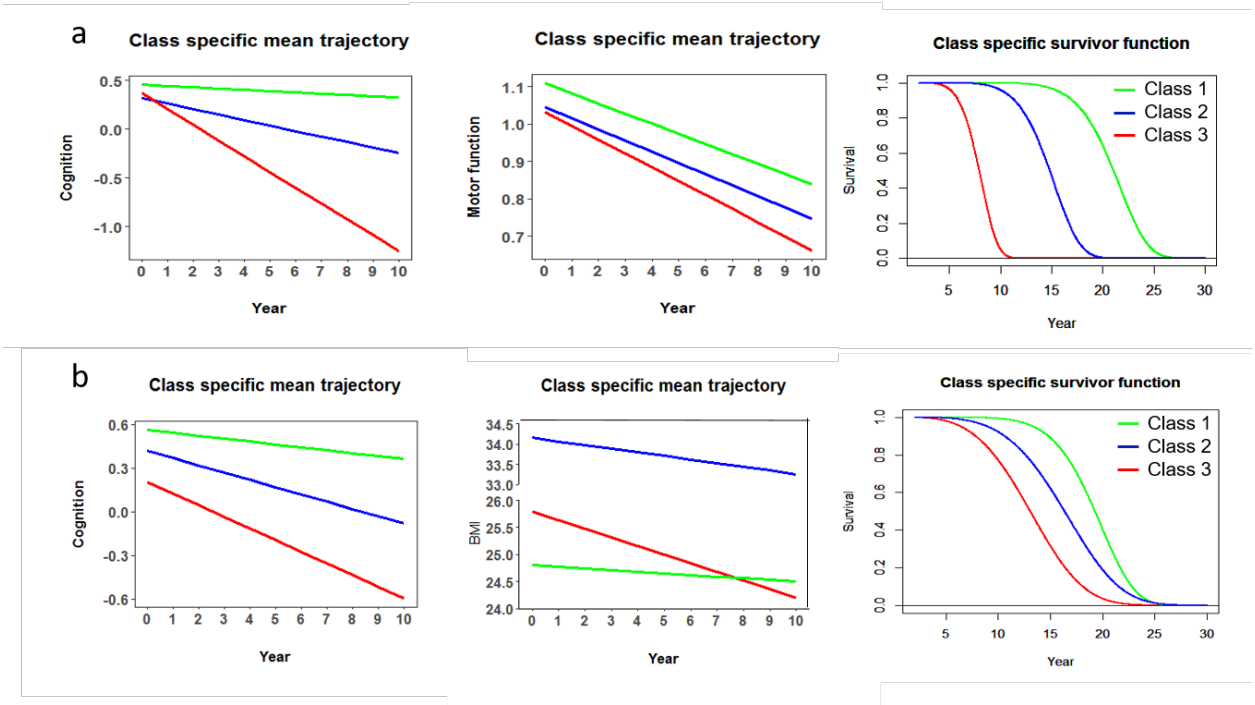

Figure 3: Predicted class-specific longitudinal trajectories and AD-free probability function a. Model of cognition and motor b. Model of cognition and BMI

sults are pretty similar. Also, we notice that as lots of observations are available, the label switching problem barely occurs.

3.3.4 Discussion

In this section, we proposed a Bayesian joint latent class model with two longitudinal outcomes and one event and applied it to MAP data. Latent classes provide a useful way for representing heterogeneity in the data. Assuming the association is explained by the latent population heterogeneity, JLCM allows us to distinguish different latent homogeneous sub-groups and to describe the corresponding class-specific profiles of aging related declines and risk of AD. Since JLCM does not have association parameters, there is no straightfor-

ward interpretation about the association. In contrast to the shared parameter joint model, differences in event risk are not explained by between-individual variation in the longitudinal response trajectories, but rather are explained by the between-class differences in the longitudinal marker profiles.

Bayesian estimation framework has been successfully applied to latent class modeling (LCM) (White and Murphy, 2014). However, Bayesian estimation has rarely been used for the joint model conditional on latent class. The estimation is mainly relying on maximum likelihood method. The difficulty is that the models need be repeatedly estimated with different initial values to avoid local maxima for each number of latent classes. Optimum number of classes is not guaranteed even after going back and forth in fitting models with different settings. In addition, when additional information is added into the joint model, for example, adding more longitudinal outcomes or considering a competing risk scenario, the number of parameters could be doubled multiple times and computation gets more difficult. Bayesian MCMC methods provide an alternative to burdensome computation needed in the multivariate frameworks. Instead of direct maximization of likelihood function, sampling from the full conditional distribution of each parameter can be reasonably straightforward. However, comparing the computation time, Bayesian usually needs much longer time to converge. Moreover, we assume a weakly informative prior for the class membership probability. If a good deal of knowledge is present about the number of latent classes, their sizes, or other parameter values, the researcher can specify a prior with a lot of certainty (or information) about these elements in the model.

The difficulty of dealing with latent class models is to determine the number and size of the classes. The reason we did not go up with 4 latent classes in the application is because model starts to show ”empty” class in the model with more than 4 classes and MCMC has difficulty to converge. A study by Nasserinejad et al., 2017 showed that in mixture model, given the prior on the class membership is sufficiently non-informative, when the model

is over fitted with more latent classes than those in the sample, the superfluous class will asymptotically become empty. The model with non-empty classes is preferred with optimal number of latent classes. Very recent work from Andrinopoulo et al,. 2020, using data in the Cystic Fibrosis cohort indicated that a class is assumed empty if it contains less than 10% of the subjects. Since our model has two longitudinal and one event outcomes, it is not practice to have too many classes which cause the interpretation and comparison between the latent classes becoming difficult. We decide to take the model with three classes for our analysis. The drawback is that when a small size of homogeneous group does exist in the data e.g. 5%, Bayesian estimation does not suit the latent class analysis in such data. This is a limitation of using Bayesian approach in our analysis for JLCM and we hope in future we can solve this problem and analyze data in which a tiny-sized group does exist.

Another difficulty when applying JLCM is that the data collected in longitudinal cohort studies may not always support the classification. The classification depends on the profiles of all the outcomes as we see the model of $Cognition\&Motor$ and model of $Cognition\&BMI$ group the MAP subjects differently. Sometimes, a clear classification may not be identifiable. When the relationship between the longitudinal trajectories and time to event is too complicated (e.g. interaction, high order intersection), interpretable numbers of classes would not satisfy the independent assumption. This is a limitation of the latent class model. This problem might be solved by adding one or more equality constraints to the model, as suggested by substantive considerations (Formann, 2011). Sometimes, the unidentified classification is related to the data validity. For example, results in this chapter are based on the sample restricted with subjects having more than 4 years of follow-up. When we choose a loose inclusion criteria to include subjects less than 4 years, we had a hard time to find theoretically meaningful and interpretable classes in which sizes of each class should not be too small or too big and independent assumption is barely held. This may be because too few observations or too short follow-up time can not sufficiently describe the longitudinal marker trajectories before the event takes place. Therefore, the relationship between trajectory shape of cog-

nitive decline and risk of AD is not clearly distinguished. Missing information may cause problems in constituting a separate class of individuals or misclassification. High quality and representative data will have more change to yield a right classification. In the other side, the identified classes may not necessarily always (and without further validation) refer to existing subgroups within the population. There could be a potential problem because, when subjects are badly separated, the classification is somewhat arbitrary, a researcher may mistakenly accept results of latent class as the solution. While dealing with these difficulties, a combination of a deep understanding of the biological and clinical mechanism, and the use of different analytic methods may provide insights into the reliability of the statistical inferences.

3.4 Alternative Approach: Joint Model with Shared Random Effects

3.4.1 Methodology

The most common modelling approach in the joint modelling literature is with shared parameter model. This type of model is composed of three components: (i) a submodel for time to event, (ii) a submodel for the longitudinal marker trajectory, and (iii) parameterization to associate (i) and (ii) processes. The basic joint model is written as

$$\begin{aligned} Y_{ij} &= x_{ij}^{\top}\beta + z_{ij}^{\top}b_i + \epsilon_{ij}, \\ h_i(t|w_i, x_i) &= h_0(t)exp(w_i^{\top}\alpha + f\{\mu_i(t), b_i, r\}),\ t > 0. \end{aligned} \tag{3.14}$$

Components in the longitudinal submodel of Y have the same interpretations as they have in the separate model (formula 3.1). In the survival submodel, w_i is a vector of covariates, possibly time-dependent, with associated vector of α representing the effects of fixed

covariates. The linking function $f(\cdot)$ have various options which lead to different forms of association between the longitudinal and time-to-event data. One is called "current value" association ($r\mu_i(t)$) in which longitudinal measure $\mu_i(t)$ is predictive of the event risk at the same time t. Second one is called "current value plus slope" association ($r_1\mu_i(t) + r_2\mu_i'(t)$). This extends the first structure by adding the slope term at time t. The third one is called "shared random effects" association ($r^\top b_i$), which includes only the random effects from the longitudinal submodel. b_i, can be viewed as the expected subject-specific trend of the longitudinal marker after adjusting for the overall mean trajectories and other covariate effects. r quantifies the association between the longitudinal outcome to the risk of an event. The longitudinal and survival modes are then considered jointly for less bias and more accurate inferences. The third type so called shared random effect model (SREM) is the most popular approach and has been extended in different ways, such as, to allow for multiple longitudinal markers, and for competing risks setting.

When extending to additional longitudinal outcome, we have

$$\begin{aligned} Y_{1ij} &= x_{1ij}^\top \beta_1 + z_{ij}^\top b_{1i} + \epsilon_{1ij}, \\ Y_{2ij} &= x_{2ij}^\top \beta_2 + z_{ij}^\top b_{2i} + \epsilon_{2ij}, \end{aligned} \tag{3.15}$$

To build the correlation between the two longitudinal processes, a multivariate normal distribution is assumed for the two sets of random effects b_i, that is:

$$b_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} \Sigma_{b1} & \Sigma_{b12} \\ \Sigma_{b12}^\top & \Sigma_{b2} \end{pmatrix} \right), \tag{3.16}$$

in which Σ_{b12} contains the covariances of the two sets of random effects. With $b_{1i} = (b_{10i}, b_{11i})$ and $b_{2i} = (b_{20i}, b_{21i})$ denoting the random intercept and slope effects in submodel of Y_1 and

Y_2, respectively, the subject-based version for the joint variance distribution is shown as:

$$\Sigma_{b1} = \begin{pmatrix} \sigma^2_{10} & Cov_{10\_11} \\ Cov_{10\_11} & \sigma^2_{11} \end{pmatrix}, \Sigma_{b2} = \begin{pmatrix} \sigma^2_{20} & Cov_{20\_21} \\ Cov_{20\_21} & \sigma^2_{21} \end{pmatrix}, \Sigma_{b12} = \begin{pmatrix} Cov_{10\_20} & Cov_{10\_21} \\ Cov_{11\_20} & Cov_{11\_21} \end{pmatrix}, \tag{3.17}$$

where $\sigma^2_{10}, \sigma^2_{11}$ and $\sigma^2_{20}, \sigma^2_{21}$ are the variances for random effects b_{10}, b_{11} and random effects b_{20}, b_{21}, respectively. Cov_{10_11} and Cov_{20_21} are the covariance between the intercept and slope, and Cov_{10_20}, Cov_{10_21}, Cov_{11_20}, and Cov_{11_21} are the covariance between the two sets of random effects.

The error terms assume following bivariate normal distribution and independent of random effects,

$$\begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} \sigma^2_1 & 0 \\ 0 & \sigma^2_2 \end{pmatrix}\right). \tag{3.18}$$

In the survival submodel, we have additional linking element $r_2^\top b_{2i}$ from the longitudinal model of Y_2.

$$h_i(t) = h_0(t) exp(\alpha^\top x_i + r_1^\top b_{1i} + r_2^\top b_{2i}), \tag{3.19}$$

where $r_1(r_{11}, r_{12})$ and $r_2(r_{21}, r_{22})$ are the coefficients linking the longitudinal and survival parts through the random effects elements. Specifically, parameters r_{11} and r_{12} denote the strength and direction of the association between Y_1 and event risk induced by random intercept and slope at the event time, respectively. r_{21} and r_{22} measure the association of Y_2 and event risk in the same way.

3.4.1.1 Estimation Method

The time to event (T_i) is conditionally independent to the longitudinal measurements (Y_{1i}, Y_{2i}) given the shared random effects b_i in the joint model. Let θ denote the combined vector of all remaining unknown parameters. Conditional on b_i, the joint log-likelihood can be written as

$$\begin{aligned} log p(T_i, \delta_i, Y_{1i}, Y_{2i}; \theta) &= log \int p(T_i, \delta_i, Y_{1i}, Y_{2i}, b_i; \theta) db_i \\ &= log \int p(T_i, \delta_i | b_i; \theta) [\prod_j p(Y_{1i}, Y_{2i} | b_i; \theta)] p(b_i; \theta) db_i, \end{aligned} \tag{3.20}$$

where the contribution to the likelihood from the event submodel is

$$p(T_i, \delta_i | b_i; \theta) = h_i(T_i | b_i, \theta)^{d_i} exp(- \int_0^{T_i} h_i(s | b_i, \theta) ds). \tag{3.21}$$

The estimates can be obtained by maximizing the marginal likelihood using the adaptive Gaussian quadrature method (Pinheiro and Bates, 1995). The derivatives of the adaptive Gaussian quadrature approximation can be computed using SAS PROC NLMIXED which performs Newton-Raphson optimization for empirical Bayes minimization of random effects.

Comparing to the maximum likelihood approach which provides only the point estimates and associated asymptotic standard error estimates for the parameters, Bayesian approach permits full and exacts posterior inference for any parameter and can also reduce computational burdens especially with a high dimension random effect structure. The inference is based on the joint posterior probability distribution of the longitudinal and survival outcomes specified in form of

$$
\begin{aligned}
p(\theta, b_i|Y_{1i}, Y_{2i}, T_i, \delta_i) &= \frac{\prod_i p(Y_{1i}, Y_{2i}|\theta, b_i)p(T_i, \delta_i|\theta, b_i)p(\theta, b_i)p(\theta)}{\prod_i p(Y_{1i}, Y_{2i}, T_i, \delta_i)} \\
&\propto \prod_i^n \{p(Y_{1i}, Y_{2i}|\theta, b_i)p(T_i, \delta_i|\theta, b_i)p(\theta, b_i)\}p(\theta)
\end{aligned} \tag{3.22}
$$

In the Bayesian approach, priors are chosen to be very vague in order to have minor impact on the posterior inferences. The standard weakly informative priors are as follows:

$\beta s \sim N(0, 1000)$,

$\alpha s \sim N(0, 1000)$,

$\sigma_\epsilon^2 \sim gamma^{-1}(0.001, 0.001)$,

$b_i \sim Wishart^{-1}(n, R)$, where $R = diag(0.01, n)$ (n=2 for univariate longitudinal outcome; n=4 for bivariate longitudinal outcomes),

$\gamma_{Weib} \sim gamma^{-1}(0.001, 0.001)$,

$rs \sim N(0, 1000)$.

Samples are drawn from the posterior distribution using the Gibbs sampler. Convergence of the sampler is checked using standard convergence diagnostics, e.g. Gelman-Rubin test. Posterior means and standard deviations of the parameters are computed from these samples and summarized.

3.4.2 Application to Memory and Aging Project Cohort

In this section, the proposed shared random effects joint model is fitted to the Rush MAP data that described previously. Our focus is to determine whether and how the cognitive performance and other related functions measured annually are associated with the time to AD dementia.

3.4.2.1 Joint Modeling of Cognitive Decline and AD Risk

We initialize the analysis with jointly modeling the cognitive decline and risk of AD. The population mean response of *Globcog* is set as $x_{ij}^{\top}\beta_{ij} = \beta_0 + \beta_1 \times time_{ij} + \beta_2 \times age\_bl_i + \beta_3 \times gender_i + \beta_4 \times edu_i + \beta_5 \times ApoE4_i$; random intercept and slope at subject specific level are $z_i^{\top} b_i = b_{0i} + b_{1i} \times time_{ij}$; the regression equation for the log-relative-hazard of AD is set as $x_{ij}^{\top}\alpha = \alpha_0 + \alpha_1 \times age\_bl_i + \alpha_2 \times gender_i + \alpha_3 \times edu_i + \alpha_4 \times ApoE4_i$. The association between cognitive decline and AD risk is through a latent zero-mean bivariate Gaussian process of random effects b_{0i} and b_{1i}, which are independent across different subjects.

Bayesian inferences is adopted and we use standard weakly informative prior distributions for the parameters. We compare 4 joint models with different forms of the latent processes $b_i^{\top} r$ linking the longitudinal and survival modes. The construction information and WAIC for each model are shown in Table VII. We start with a simple model I without random effects and the WAIC (631.8) is large indicating a poor fit. In Model II, we first introduce association through the random intercept b_0, which leads to a substantial decreased WAIC (615.4) of the joint model, suggesting an association between the two submodels. Then we further allow both random intercept and slope in the association between the two submodels (Model III and Model IV). Comparing the WAIC of all the models, model IV yields the smallest value (274.9), indicating the best fit. Under this model, r_1 and r_2 are different. The result is consistent with what is found in BIC when we apply maximum likelihood estimation.

Table VII: MODEL SELECTION FOR JOINT MODELS WITH DIFFERENT ASSOCIATION PATTERNS

Model	$b_i^{\top} r$	WAIC(Bayesian)	BIC(MLE)
I	0	631.8	3441
II	$r_1 b_0$	615.4	3418
III	$r_1(b_0 + b_1)$	495.2	3389
IV	$r_1 b_0 + r_2 b_1$	274.9	3120

We present in Table VIII (right) the posterior means of the interested parameters with 95% credible intervals based on Model IV. We found the results are similar to what we get from the maximum likelihood approach shown in Table VI left. The association between cognitive decline and AD incidence is explained by parameter r_1 (-0.326) and r_2 (-5.719), indicating that a subject's risk of AD is related to the cognition level at entry and the cognitive decline rate. This is clinically reasonable, since better cognitive performance represents better mental status; subjects with more rapid cognitive decline would be expected to have a higher risk of AD. In the joint model, time, age at baseline, gender, education level but ApoE4 are statistically significant with the cognition. In the hazard submodel, only age at baseline is positively associated with the risk of AD and the effect of age decreases about a half of that in the separate survival model. ApoE4, which shows a strong association with AD risk in the separate hazard model (Table I), has no effect on AD incidence in the joint model. The different results from the joint model tells us that information from cognition decline has an impact on the estimation of time to AD dementia.

We show by the joint model the strong association between the cognitive decline and AD incident through random effects. To further investigate this finding, we plot the estimated posterior density (Figure 4a) of the median AD onset time for a participant from our study. This subject did not have AD diagnosed during the observation period. The prediction from the separate survival model (red) is about 10 to 20 years (Peak is 15 years) after the first visit. Survival model with cognition score as a time-varying covariate (blue) decreases the estimated AD onset time by approximately 5 years compared with the separate model. The plot based on the joint modeling (green) is closer to the plot from the survival model with time-varying variable for cognition, but predicts a little later AD onset time. We show in Figure 4b this subject's specific cognition trajectory. Comparing to the predicted cognitive decline rate for overall cohort (orange), this subject has a relatively "bad" trajectory of cognitive decline (Green line is the subject-specific profile predicted from joint model.). The selected subject has characteristics associated with a poor brain function degradation and

Table VIII: PARAMETER ESTIMATES OF SHARED RANDOM EFFECTS MODEL ON COGNITIVE DECLINE AND AD RISK

	MLE		Bayesian	
Parameter	Mean	95%CI	Posterior mean	95%CI
Longitudinal				
Intercept (β_0)	0.927	(0.580, 1.273)	0.943	(0.589, 1.258)
Time (β_1)	-0.048	(-0.053, -0.043)	-0.048	(-0.052, -0.042)
Age_fb (β_2)	-0.017	(-0.021, -0.013)	-0.017	(-0.021, -0.013)
Gender (β_3)	-0.070	(-0.132, -0.009)	-0.071	(-0.133, -0.009)
Edu (β_4)	0.054	(0.046, 0.063)	0.055	(0.047, 0.062)
ApoE4 (β_5)	0.041	(-0.023, 0.106)	0.043	(-0.017, 0.107)
Ran_int (σ_0^2)	0.106	(0.093, 0.119)	0.107	(0.095, 0.121)
Ran_Cov (Cov_{01})	-0.000	(-0.002, 0.002)	-0.000	(-0.002, 0.002)
Ran_time (σ_1^2)	0.004	(0.003, 0.004)	0.004	(0.003, 0.004)
Error (σ_ϵ^2)	0.049	(0.047, 0.050)	0.049	(0.047, 0.050)
Proportional hazard				
Intercept (α_0)	-4.675	(-5.218, -4.131)	-4.589	(-4.777, -4.409)
Age_fb (α_1)	0.022	(0.016, 0.028)	0.022	(0.020, 0.024)
Gender (α_2)	0.059	(-0.022, 0.140)	0.049	(-0.025, 0.119)
Edu (α_3)	0.002	(-0.009, 0.013)	0.002	(-0.010, 0.007)
ApoE4 (α_4)	-0.015	(-0.090, 0.059)	-0.012	(-0.075, 0.059)
Linking				
Ran_int (r_1)	-0.325	(-0.413, -0.236)	-0.326	(-0.413, -0.235)
Ran_time (r_2)	-5.958	(-6.635, -5.282)	-5.719	(-6.424, -5.136)

this is reflected a shorter predicted time to AD onset in the joint model. Separate survival analysis purely based on the demographic and genotype provides a late onset time, which is very biased. The difference in the predicted onset time is explained by the fact that the joint model correctly accounts for the correlation between the cognitive function and the risk of AD. Models with time-varying covariates can help to adjust the bias as shown in the figure, but can also lead to an overestimated or underestimated result.

3.4.2.2 Joint Modeling of Cognition and Motor Function Decline

We next investigate the two longitudinal outcomes (cognition and motor) jointly in a mixed model with a shared random effects structure. The model is fitted using Bayesian

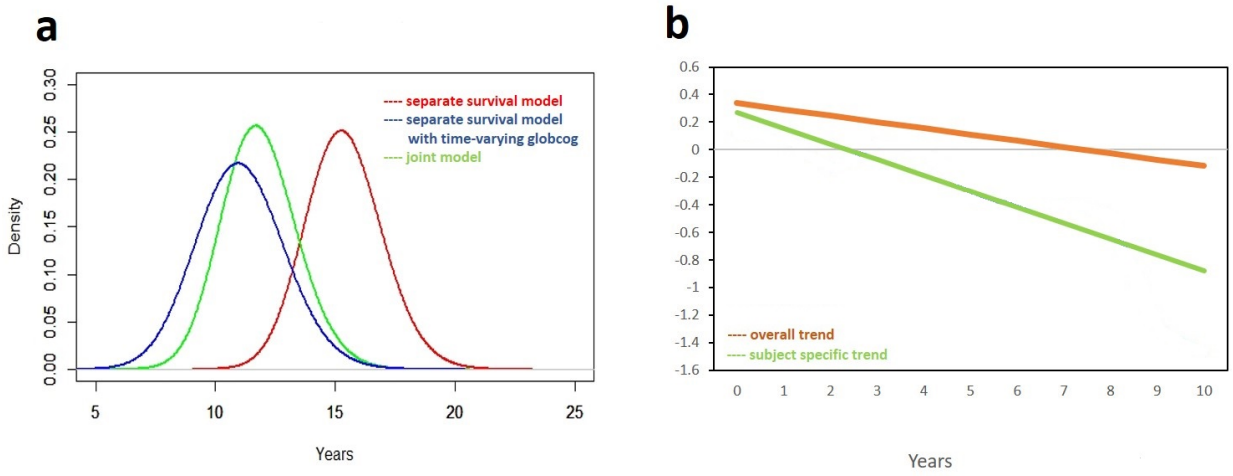

Figure 4: Posterior density of time to AD and trajectory of cognition for a participant

approach. Particularly, we take inverse-Wishart prior with 4 degrees of freedom for the variance-covariance matrix of the random effects. Compared to the model without the covariances between the two sets of random effects, WAIC is reduced from -7005.1 to -7067.5 for the proposed model. The results are shown in Table IX. Both global cognitive function and motor function scores significantly decrease across time with $\beta = -0.048$ (95%CL: -0.052, -0.043) and $\beta = -0.030$ (95%CL: -0.031, -0.029), respectively. Age from baseline, gender, education are associated with cognition and motor function. BMI has a negative effect on motor function. The covariances between the two sets of the random effects Cov_{10_20} and Cov_{11_21} are positive with 95% credible intervals different from 0, suggesting a significant association. The individual level random effects correlation between the two outcomes indicates the motor dysfunction and cognitive decline tend to go hand by hand as AD progresses and should be modelled jointly.

Table IX: PARAMETER ESTIMATES OF JOINT MODEL FOR COGNITION AND MOTOR FUNCTION DECLINE

Parameter	Posterior mean	95%CL	Parameter	Posterior mean	95%CL
Outcome=Globcog			Outcome=Globmot		
Intercept (β_{10})	0.981	(0.630, 1.317)	Intercept (β_{20})	2.676	(2.502, 2.848)
Time (β_{11})	-0.048	(-0.052,-0.043)	Time (β_{21})	-0.030	(-0.031, -0.029)
Age_bl (β_{12})	-0.017	(-0.021, -0.013)	Age_bl (β_{22})	-0.018	(-0.020, -0.016)
Gender (β_{13})	-0.073	(-0.138, -0.014)	Gender (β_{23})	0.063	(0.037, 0.087)
Edu (β_{14})	0.054	(0.044, 0.062)	Edu (β_{24})	0.005	(0.002, 0.009)
ApoE4 (β_{15})	0.030	(-0.030, 0.099)	BMI (β_{25})	-0.010	(-0.012, -0.008)
Ran_int (σ^2_{10})	0.108	(0.096, 0.122)	Ran_int (σ^2_{20})	0.024	(0.021, 0.027)
Ran_time (σ^2_{11})	0.003	(0.003, 0.004)	Ran_time (σ^2_{21})	0.000	(0.000, 0.000)
Ran_cov (Cov_{10_11})	-0.000	(-0.002, 0.002)	Ran_cov (Cov_{20_21})	-0.001	(-0.001, -0.000)
Error ($\sigma^2_{\epsilon 1}$)	0.048	(0.047, 0.050)	Error ($\sigma^2_{\epsilon 2}$)	0.013	(0.013, 0.014)
Ran_cov (Cov_{10_20})	0.011	(0.007, 0.016)			
Ran_cov (Cov_{10_21})	0.001	(0.001, 0.000)			
Ran_cov (Cov_{11_20})	0.000	(-0.002, 0.002)			
Ran_cov (Cov_{11_21})	0.000	(0.000, 0.000)			

3.4.2.3 Joint Model of Cognition and Motor Decline and AD risk

It will be interesting to include both cognition and motor function when studying the subjective aging trajectories and AD risk in a joint modeling framework. In application, this will be more informative because there should be more than one potential factor or bio-marker related to disease.

Under the shared random effects modelling framework, we assume that there is a stochastic dependence between cognition and motor submodels and the risk of AD through the random effects. Based on the result from the above section, we would like to take account for the subject level correlation between cognition and motor by assuming a multivariate normal distribution for their random effects (b_i in equation 3.16). For the survival part, we postulate the linking process by $b_i^\top r = r_{11}b_{10i} + r_{12}b_{11i} + r_{21}b_{20i} + r_{22}b_{21i}$, where (r_{11}, r_{12}) and (r_{21}, r_{22}) are the linking parameters for the random intercept and slope effects of cognition and motor function, respectively. Next, we perform model selection among models with different asso-

ciation patterns. Model with fewer number of parameters and lower WAIC is preferred. We allow a significantly longer burn-in period for each model followed by convergence diagnostics in WinBUGS.

Table X shows the WAIC for models with different association patterns for $b_i^\top r$. Comparing to Model I (WAIC=-8001.9) without shared random effects, the model fit improves with either adding the random effects from longitudinal model of cognition (Model III, WAIC=-8376.8) or of motor (Model V, WAIC=-8128.1). However, when cognitive model and motor model both share random effects with the AD risk model (Model VI-IX), WAIC does not decrease as low as that of Model III. As we like to keep the model simple, Model III with fewer number parameters and the smallest WAIC is preferred. It seems that the random effects from the cognition model are sufficient to account for the association between the longitudinal trajectories and AD incidence. We also evaluate the Model III* which is the same as Model III except for an independent distribution of the cognition and motor trajectories. Higher WAIC (-8337.8) for Model III* compared to Model III suggests the correlation of the two sets of random effects must be modelled in the joint model and the association of cognitive decline and AD risk to be investigated need the adjustment of motor function.

Table XI shows the posterior means and 95% credible interval for Model III using Bayesian estimation procedure. The estimated association parameters rs ($\gamma_{11} = -0.332$, $\gamma_{12} = -5.806$) are negative and credible interval do not cover 0, meaning that higher initial level of cognitive performance and slower cognitive decline rate are associated with a lower risk of AD. In the random effects variance-covariance matrix (4×4) of the cognition and motor longitudinal submodels, elements Cov_{10_20} and Cov_{11_21} reflecting the correlation of the random intercepts and slopes between the two longitudinal models both have positive values significantly different from zero.

Table X: MODEL SELECTION FOR JOINT MODELS WITH DIFFERENT ASSOCIATION PATTERNS

Model	$b_i^\top r$	WAIC
No shared random effect		
I	0	-8001.9
Random effects from model of cognition		
II	$r_{11}b_{10}$	-8020.5
III	$r_{11}b_{10} + r_{12}b_{11}$	-8376.8
III*	$r_{11}b_{10} + r_{12}b_{11}$	-8337.8
Random effects from model of motor		
IV	$r_{21}b_{20}$	-8019.4
V	$r_{21}b_{20} + r_{22}b_{21}$	-8128.1
Random effects from both		
VI	$r_{11}b_{10} + r_{21}b_{20}$	-8006.8
VII	$r_{11}b_{10} + r_{12}b_{11} + r_{21}b_{20}$	-8358.8
VIII	$r_{11}b_{10} + r_{21}b_{20} + r_{22}b_{21}$	-8155.6
IX	$r_{11}b_{10} + r_{12}b_{11} + r_{21}b_{20} + r_{22}b_{21}$	-8358.9

* The random effects of the cognition and motor are set as independent.

To further investigate the predictive quality on survival time, we compare the estimated posterior density of the median AD onset time based on Model III and Mode III*. We present in Figure 5 the estimated subject-specific median AD onset time for a participant without diagnosed AD in MAP cohort. The plot from Model III (green) differs from Model III* (red), which does not take account the correlation between the cognition and motor function trajectories, slightly decreasing the onset time about 2-3 years. The difference is due to the adjustment of the motor function decline which helps to gain more power in the analysis by inducing more accurate longitudinal bio-mark information to the event time model.

Table XI: PARAMETER ESTIMATES OF JOINT MODEL FOR COGNITION AND MOTOR DECLINE WITH AD RISK

Parameter	Posterior mean	95%CL	Parameter	Posterior mean	95%CL
Outcome=Globcog			Outcome=Globmot		
Intercept (β_{10})	0.874	(0.523, 1.210)	Intercept (β_{20})	2.667	(2.497, 2.842)
Time (β_{11})	-0.049	(-0.053, -0.044)	Time (β_{21})	-0.030	(-0.031, -0.029)
Age_bl (β_{12})	-0.016	(-0.020, -0.012)	Age_bl (β_{22})	-0.018	(-0.019, -0.016)
Gender (β_{13})	-0.075	(-0.146, -0.018)	Gender (β_{23})	0.064	(0.039, 0.090)
Edu (β_{14})	0.055	(0.047, 0.064)	Edu (β_{24})	0.006	(0.002, 0.009)
ApoE4 (β_{15})	0.033	(-0.031, 0.099)	BMI (β_{25})	-0.010	(-.012, -0.008)
Ran_int (σ^2_{10})	0.107	(0.095, 0.120)	Ran_int (σ^2_{20})	0.004	(0.003, 0.004)
Ran_time (σ^2_{11})	0.024	(0.021, 0.027)	Ran_time (σ^2_{21})	0.000	(0.000, 0.000)
Ran_cov (Cov_{10_11})	-0.000	(-0.002, 0.002)	Ran_cov (Cov_{20_21})	-0.001	(-0.001, 0.000)
Error ($\sigma^2_{\epsilon_1}$)	0.049	(0.047, 0.050)	Error ($\sigma^2_{\epsilon_2}$)	0.013	(0.013, 0.014)
Ran_cov (Cov_{10_20})	0.011	(0.007, 0.016)			
Ran_cov (Cov_{10_21})	-0.001	(-0.001, -0.000)			
Ran_cov (Cov_{11_20})	0.001	(0.000, 0.002)			
Ran_cov (Cov_{11_21})	0.000	(0.000, 0.000)			
Outcome=time to AD					
Intercept (α_0)	-4.237	(-4.547, -4.036)			
Agefb (α_1)	0.017	(0.015, 0.020)			
Gender (α_2)	0.064	(-0.012, 0.139)			
Edu (α_3)	-0.002	(-0.010, 0.007)			
ApoE4 (α_4)	-0.011	(-0.087, 0.065)			
Association parameters					
Ran_int (r_{11})	-0.332	(-0.416, -0.203)			
Ran_time (r_{12})	-5.806	(-6.506, -5.221)			

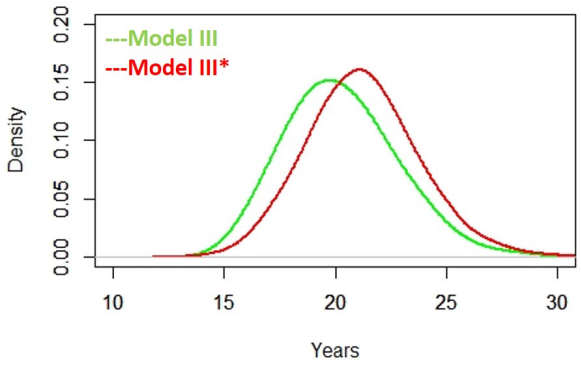

Figure 5: Posterior distribution of median AD onset time in Model III and Model III*

3.4.3 Discussion

In this section, we use a joint modeling approach through shared random intercept and slope effects to analyze the longitudinal marker and time to event simultaneously. We show the joint model provides a better fit to the data and a better prediction for the risk of AD. Bayesian and maximum likelihood methods are both used and they provide similar results. Comparing Bayesian versus maximum likelihood approach, Bayesian method is relatively straightforward to implement with simpler coding and has comparable running time when model is simple. In addition, we notice the local maximum problem with Newton-Raphson optimization and need to repeat the estimation with different sets of initial values.

Most joint modeling analysis are typically dealing with a single longitudinal outcome and a primary event. Motivated by the clinical interest, we proposed an extended joint model which handles bivariate longitudinal outcomes and time-to-AD. In practice, for the purpose of medical decision-making, clinical studies are likely to gather all possible informative inferences and incorporate as many as sources of data to improve prediction.

In this chapter, two major joint modeling frameworks, latent class and shared random effects, are discussed. Both approaches improve the model fitting and accuracy of predictions of event risk. However, the former paradigm has no clear interpretation for the association of the longitudinal and survival outcomes while the latter dose not account for the heterogeneity. Recently, an interesting work done by Andrinopoulo et al., 2020, incorporated the latent classes into the shared random effects model for a heterogeneous population. Their proposed model is able to assess the association between the longitudinal and survival processes while allowing for latent classes but requires an intensive computational effort.

4 JOINT MODEL WITH RANDOM CHANGEPOINT IN LONGITUDINAL MARKER TRAJECTORY AND COMPETING RISKS IN SURVIVAL PROCESS

4.1 Trajectories of Longitudinal Marker with a Random Changepoint

Sometimes, a linear random effects model may not be suitable in the longitudinal analysis, if the trajectories of longitudinal marker shows nonlinearity over time. For instance, AD is characterized by a very long-diagnosis decline, based on literature, we know that an acceleration of the cognitive decline occurs about 2 to 5 years before the diagnosis of AD (Amieva et al., 2005). Figure 6 shows state 0 (healthy) to state 1 with slight linear decline in cognition corresponding to normative age-graded influences and state 1 to state 2 with an accelerated decline corresponding to the pre-diagnosis phase of AD. Such data exhibiting a trend of changing in direction are commonly observed in medical and environmental science, and the two approximately linear phases are assumed linked with continuous transition with a changepoint. Understanding the shape of this change and the critical time point for subjects who develop event is a challenge and becomes increasingly important for early detection of subjects at high risk of disease.

Previous studies modeled the trajectories using a quadratic time function or spline function, which does not allow the identification of the transition from phase 1 to phase 2 for an accelerated drop rate. Changepoint models have been used to describe the trajectories and trend of longitudinal measures and allow different linear functions of time corresponding to the pre- and post-critical time point trends. In the history of the changepoint model, early in 1970, Hinkley first considered inference about the changepoint problem in a sequence of random variables from a frequentist approach (Hinkley, 1970). The idea was later extended in a Bayesian framework for a continuous version of the changepoint (Carlin et al., 1992. In 2000, Hall first used a piecewise linear mixed model to compare between AD cases vs

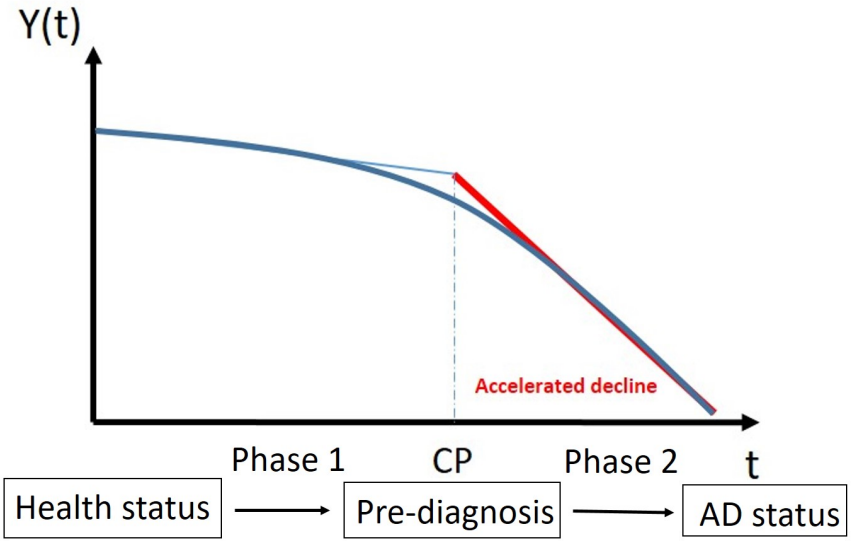

Figure 6: Trajectory of change in global cognition with before diagnosis of AD

AD free subjects the trajectory of cognitive functions (Hall et al., 2000). Later, to relax the assumption of all subjects having the same changepoint time, he applied Bayesian approach in a random changepoint model to describe the cognitive decline (Hall et al., 2003). The piecewise linear model (broken-stick) is a first-step candidate to describe two linear phases with an abrupt transition and is most widely used. In practice, such sharp change is not realistic and the non-continuity at the changepoint may cause numerical issue in parameter estimation, such as the maximum likelihood method. Therefore, a more flexible smooth changepoint model could characterize the overall natural trend of transition. The general framework of the random changepoint model can be expressed as

$$E(Y_{ij}) = \begin{cases} f_1(t_{ij}) & t_{ij}\ before\ transition_{ij} \\ f_2(t_{ij}) & t_{ij}\ during\ transition_{ij} \\ f_3(t_{ij}) & t_{ij}\ after\ transition_{ij}, \end{cases} \tag{4.1}$$

where $f_1(.)$ and $f_3(.)$ are functions for the two linear phases and $f_2(.)$ is a chosen function that describes the trend in the transition zone. Depending on the form of $f_2(.)$, changepoint models can be classified into different types. Next, five changepoint models for multiphase longitudinal data will be discussed: the piecewise (broken-stick) model, Bacon-Watts model (BW), Griffiths-Miller model (GM), bent-cable model (BC) and polynomial regression model (PR). All the models introduced are represented in Figure 7.

4.1.1 Broken-Stick Model

For the observed outcome Y_{ij} given t_{ij} and changepoint τ_{ij} , the broken-stick model is given by

$$E(Y_{ij}) = \begin{cases} \beta_0 + \beta_1 t_{ij} & t_{ij} \leq \tau_i \\ \beta_0 + \beta_1 t_{ij} + \beta_2 (t_{ij} - \tau_i) & t_{ij} > \tau_i \end{cases} \tag{4.2}$$

where β_0 and β_1 are, respectively, the intercept and slope before the changepoint τ , and $\beta_1 + \beta_2$ is the slope after the changepoint. The broken-stick model with sudden change in direction can be implemented in a Bayesian framework and parameter interpretation is relatively easy.

4.1.2 Bacon-Watts Model

An smooth changepoint model introduced by Bacon and Watts, 1971, is given as

$$E(Y_{ij}) = \beta_0 + \beta_1 (t_{ij} - \tau_i) + \beta_2 (t_{ij} - \tau_i) trn((t_{ij} - \tau_i)/\gamma), \tag{4.3}$$

where β_0 is the value of outcome at the changepoint τ. trn is the transition function which smooths the intersection between the two lines and γ is a transition parameter. The function of trn is restricted with the following assumptions:

1. $\lim_{t\to\infty} trn(|t|/\gamma) = 1$
2. $trn(0) = 0$
3. $\lim_{\gamma\to 0} trn(|t|/\gamma) = sgn(t)$
4. $\lim_{t\to\infty} strn(|t|/\gamma) = t$

Large value of the transition parameter γ indicates a gradual transition. In particular, if γ is close to zero, the Bacon-Watts model will imply a quick change. The Bacon-Watts model implies a smooth change, however, it has some shortcomings. One disadvantage is that the interpretations of β_1 and β_2 are not straightforward and no longer have the same meaning as the broken-stick model. Another problem is that a slight increase of the slope in decline right before transition is implied in the model, which is not realistic for the sample with monotonic change in rate, for example, cognitive decline in dementia applications.

4.1.3 Griffiths-Miller Model

In order to avoid the bulge just before the changepoint, Griffiths and Miller, 1973, based on the Bacon-Watts model, excluded above assumption 2 and used $trn(t) = \sqrt{t^2+\gamma}$ as a transition function (Griffiths and Miller, 1973). They proposed

$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2\sqrt{(t_{ij}-\tau_i)^2+\gamma}, \tag{4.4}$$

where β_0 denotes the value of outcome at the changepoint. Direct interpretations of β_1 and β_2 are not available. Only when the γ is small enough, $\beta_1 - \beta_2$ is close to the slope before the changepoint and $\beta_1 + \beta_2$ is the slope after.

4.1.4 Bent-Cable Model

Chiu et al., 2006, considered a regression framework, so called bent-cable to analyze data exhibit smooth transition. The bent-cable methodology provides sufficient flexibility and interpretability. The model is parsimonious and can be expressed as

$$E(Y_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 q(t_{ij}; \tau_i; \gamma), \tag{4.5}$$

where

$$q(t_{ij}; \tau_i; \gamma) = \frac{(t_{ij} - \tau_i + \gamma)^2}{4\gamma} I[\tau_i - \gamma < t_{ij} \leq \tau_i + \gamma] + (t_{ij} - \tau_i) I[t_{ij} > \tau_i + \gamma], \tag{4.6}$$

in which, $I[*]$ is an indicator function that equals 1 if * is true and 0 otherwise. Like the broken-stick model, β_0 and β_1 are, respectively, the intercept and slope before the changepoint τ. The $q(.)$ function assumes a quadratic bend. γ is the transition parameter and two γs cover the transition zone with τ in the center. In other words, the transition starts at time $\tau - \gamma$ and ends at time $\tau + \gamma$. When γ is very small, this model is close to the broken-stick model. $\beta_0 + \beta_1$ is the slope of the linear phase after changepoint. Sometimes Bacon-Watts is also called the bent-cable model, which could cause confusion.

4.1.5 Polynomial Regression Model

Hout et al., 2010, proposed a model, called polynomial model, with regard to the modelling of cognitive decline to link the two linear parts with a curve using a polynomial function. The random smooth polynomial model is given by

$$E(Y_{ij}) = (\beta_0+\beta_1 t_{ij})I[t_{ij} < \tau_i]+g(t_{ij};\beta_0;\beta_1;\beta_2;\gamma)I[\tau_i \leq t_{ij} < \tau_i+\gamma]+(\lambda_i+\beta_2 t_{ij})I[t_{ij} \geq \tau_i+\gamma], \tag{4.7}$$

where g is a third-degree polynomial connecting the two lines. β_0 and β_1 are interpreted the same as in the broken-stick model. Change in slope starts at τ and ends at $\tau + \gamma$. With γ=0 the model becomes a broken-stick model. β_2 is the linear slope after change has taken place. Smoothness of the transition is implied with following constraints as:

$$\begin{aligned} g(\tau_i) &= \beta_0 + \beta_1\tau_i \\ g(\tau_i + \gamma) &= \lambda + \beta_2(\tau_i + \gamma) \\ \frac{\partial g}{\partial t_{ij}}(\tau_i) &= \beta_1 \\ \frac{\partial g}{\partial t_{ij}}(\tau_i + \gamma) &= \beta_2 \end{aligned} \tag{4.8}$$

Parameter λ is derived by assuming the continuity of the two lines at $\tau + \gamma/2$, which is defined as the changepoint. λ can be represented by a function of $\beta_0, \beta_1, \beta_2$ and γ, and based on the assumption, it is expressed as $\lambda = \beta_0 + \beta_1(\tau_i + 1/2\gamma) - \beta_2(\tau_i + 1/2\gamma)$.

The model has the advantage of having a direct parameter interpretation on both linear parts. The high dimensional transition allows smoother transition regimes. The shape of the transition is not directly determined by the first linear phase but the second, thus it does not reveal useful information regarding the rate change between the two phases.

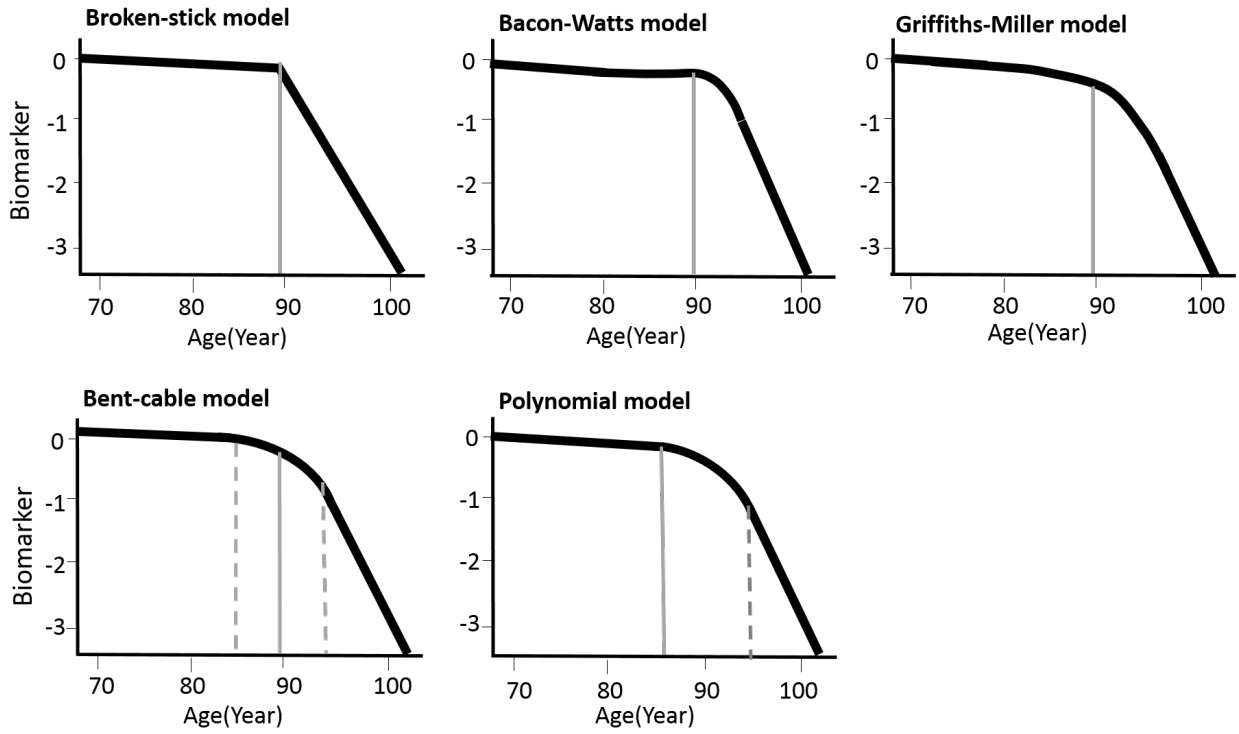

Figure 7: Estimated trajectories based on the Broken-stick model, the Bacon-Watts model, Griffiths-Miller model, bent-cable model and Polynomial model.
The grey solid line represents the value of τ; the grey dashed lines are the limits of the bend area estimated by transition parameter γ.

Since joint modeling framework offers many advantage over the separate analysis, Jacqmin-Gadda et al., 2006, combined a piecewise model for an acceleration in cognitive change with a log-normal survival model. Later, Yu and Ghosh, 2010, extended the joint model with a mixture model for survival data with death competing risk. Both groups considered the connection between the longitudinal submodel and survival submodel only depending on the log transformed random changepoint $log(\tau)$, which is used as a covariate in the proportion hazard function. No additional parameters from the longitudinal model were shared with the survival model.

4.2 Subsurvival Function with Competing Risks

In survival analysis, an important consideration is informative censoring, namely when participants without event are lost to follow-up before the end point due to reasons related to the study. For example, the prevalence of AD is high among the elderly, the follow-up in the study cohort is often interrupted by death. Participants may die before the visit following AD onset, without being diagnosed. Also AD and death are highly correlated and share common risk factors, such as age and gender. Thus the assumption that censored subjects have the same AD hazard as those at-risk is not fulfilled. Neglecting the competing risks in time-to-event will cause the estimated effect of a factor on the risk of event to be biased, particularly for elderly patients with multimorbidity.

In the standard competing setting, the outcome is the time to either of the events or the censoring time with an indicator for different events (δ=1 if event 1 happens first, δ=2 if event 2 happens first and δ=0 if censored). Then the cause-specific hazard rate function is modified from hazard function (Equation 3.3) as:

$$h_k(t) = \lim_{\Delta t \to 0} \frac{P(t <= T < t + \Delta t, \delta = k) | (T >= t)}{\Delta t}. \tag{4.9}$$

The overall survival likelihood function is a product of multiple likelihoods, one for each type of failure:

$$L = \prod_{k=1}^{2} h_j P(t <= T < t + \Delta t, \delta = k) | (T >= t) \Delta t. \tag{4.10}$$

Note that some event onset time could be interval censored (e.g. AD diagnosed time) while some event occurrence can be observed exactly (e.g. death). We assume we have both types. For the first type of event, the failure time is between two observed times t_1 and t_2 and the likelihood for interval censored cases is $L_{Event1i} = [S(t_1) - S(t_2)]^{\delta_{Event1i}}$. For the second

type of case, the likelihood is $L_{Event2_i} = f(t)^{\delta_{Event2i}}$. With right censored $L_i = S(t)^{1-\delta_i}$, the complete likelihood for survival is:

$$L_i = [S(t_1) - S(t_2)]^{\delta_{Event1i}} f(t)^{\delta_{Event2i}} S(t)^{1-\delta_i}. \tag{4.11}$$

The hazard function $h_k(t)$ is taken to be Weibull proportional hazard specification with shape (a_k) parameter and scale (λ_k) parameter as $h(t|\lambda_k, a_k)_k = \lambda_k a_k t^{(a_k-1)} exp(x^\top \alpha)$, k=1,2 for two competing events.

4.3 Joint Modeling of Longitudinal Outcomes with Changepoints and Competing Events

4.3.1 Framework of the Joint Model

Since smooth changepoint models which are more realistic in practice have never been jointly estimated with survival outcomes, we like to propose a joint model considering the above introduced random changepoint frameworks for longitudinal marker trajectories. For the typical 2-phase shaped change across time, such as cognitive decline, we consider just one changepoint trajectory that comprises two linear parts linked with a curved bend.

We apply a linear mixed effects model in which the primary time is t. Let Y_{ij} be the longitudinal response of subject i at t_j; τ_i is the time at changepoint of longitudinal response of subject i; ϵ_{ij} is the residual error and is independently distributed. Additional covariates x_i are considered as $\beta_0 = \beta_{00} + \beta_{01}^\top x_i$. Both population-level trend (β_0, β_1 and β_2) and individual-level effects (b_{0i}, b_{1i} and b_{2i}) are included for the intercept and slopes. Together we have $\beta_{0i} = \beta_0 + b_{0i}$ (intercept), $\beta_{1i} = \beta_1 + b_{1i}$ (slope before changepoint),

$\beta_{2i} = \beta_2 + b_{2i}$ (slope after changepoint, interpretation varies with different settings), where $b_i = (b_{0i}, b_{1i}, b_{2i})^\top$ is multivariate normally distributed $N(0, \Sigma_b)$ with zero means and Σ_b as a positive-definite matrix. The random changepoint τ_i is independent from random effects b_i. Based on the transition function, we have different models describing the trajectories of longitudinal marker.

In addition, we assume that the random changepoint τ_i follows a truncated normal distribution $N(\mu_{\tau_i}, \sigma^2_{\tau_i})$ with the constraint that τ_i is within a reasonable limit, for example, between age of 50 and 120 for cognition, given little decline in cognitive performance occurs until people are about 50 years old (Salthouse, 2009). Furthermore, the mean μ_{τ_i} is depending on a row vector of covariates z_i.

For survival submodel, we assume that event time follows a Weibull proportional hazard model with $S(t) = exp(\lambda a t^{a-1} exp(x^\top \alpha))$. The model will be enriched with considering the competing risk (event 1 and event 2) and interval censored time of event 1. We here define a shared random effects joint model linking the longitudinal marker trajectory and the time to events. Besides $x^\top \alpha$ in the proportional hazard function, the random changepoint τ_i and random effects of intercept and slopes b_i are included to link the two processes.

With above described main components, we define the joint model with changepoint expressed as:

Longitudinal submodel:

$$
\begin{aligned}
&Model\ (1)\ Broken-stick\\
&Y_{ij}=(\beta_{00}+\beta_{01}x_i+b_{0i})+(\beta_1+b_{1i})t_{ij}+(\beta_2+b_{2i})(t_{ij}-\tau_i)I(t_{ij}-\tau_i)+\epsilon_{ij}\\
&Model\ (2)\ Batton-Watts\\
&Y_{ij}=(\beta_{00}+\beta_{01}x_i+b_{0i})+(\beta_1+b_{1i})(t_{ij}-\tau_i)+(\beta_2+b_{2i})trn[(t_{ij}-\tau_i)/\gamma]+\epsilon_{ij}\\
&Model\ (3)\ Griffiths-Miller\\
&Y_{ij}=(\beta_{00}+\beta_{01}x_i+b_{0i})+(\beta_1+b_{1i})(t_{ij}-\tau_i)+(\beta_2+b_{2i})\sqrt{(t_{ij}-\tau_i)^2+\gamma}+\epsilon_{ij}\\
&Model\ (4)\ Bent-cable\\
&Y_{ij}=(\beta_{00}+\beta_{01}x_i+b_{0i})+(\beta_1+b_{1i})t_{ij}+\\
&(\beta_2+b_{2i})\left(\frac{(t_{ij}-\tau_i+\gamma)^2}{4\gamma}I[\tau_i-\gamma<t_{ij}\leq\tau_i+\gamma]+(t_{ij}-\tau_i)*I[t_{ij}>\tau_i+\gamma]\right)+\epsilon_{ij}\\
&Model\ (5)\ Polynomialregression\\
&Y_{ij}=[(\beta_{00}+\beta_{01}x_i+b_{0i})+(\beta_1+b_{1i})t_{ij}]I[t_{ij}<\tau_i]+g(t_{ij},\beta_0,\beta_1,\beta_3,b_{0i},b_{1i},b_{2i},\gamma)I[\tau_i\leq t_{ij}<\\
&\tau_i+\gamma]+[\lambda_i+(\beta_3+b_{3i})t_{ij}]I[t_{ij}\geq\tau_i+\gamma]+\epsilon_{ij}
\end{aligned}
\tag{4.12}
$$

Changepoint submodel:

$$\mu_{\tau_i}=\beta_{\tau_0}+\beta_{\tau_1}z_i+\epsilon_{\tau_i} \tag{4.13}$$

Survival submodel:

$$
\begin{aligned}
&h_k(t|h_{0k},\alpha_k,\zeta_k,r_k)=h_{k0}(t)exp(\alpha_k x_i+\zeta_k\tau_i+r_{1k}b_{0i}+r_{2k}b_{1i}+r_{3k}b_{2i}),\\
&where\ k=1,2.
\end{aligned}
\tag{4.14}
$$

The fixed effects βs may have different interpretations depending on the time scaling. For example, in broken-stick, bent-cable and polynomial model, β_0 notes the value of longitudinal

response at t_{ij}=0, which is not so meaningful when the time is using birth age, while β_0 in Bacon-Watts and Griffiths-Miller model is the longitudinal value at changepoint τ_i; β_1 in broken-stick, bent-cable and polynomial model is the slope before changepoint and β_2 represents the slope difference of the 2 linear phases in broken-stick and bent-cable model while β_2 is the slope of phases 2 in polynomial model. In Bacon-Watts and Griffiths-Miller, the interpretations of β_1 and β_2 are difficult because of the formulation, especially when the transition parameter γ has a large value.

Note that γ is used in the smooth changepoint model (2-5) as a transition parameter to provide flexibility to handle the transition zone (i.e. shape and width of the transition curve). The value of γ can be data driven, rather than pre-determined value. As $\gamma \to 0$, any directional change of the slope occurs at τ_i. All the smooth changepoint models are reduced to broken-stick model (1) for an abrupt transition.

This link between the longitudinal and survival model is through shared random changepoint and random effects. We assume the independence of the random changepoint and random effects in the longitudinal model. Depending on the data to be analyzed, the shared random element structure could be considered to link the longitudinal submodel with the risk of the primary event or both events in the competing risk survival model. For example, in cognition and dementia analysis, we think it is necessary to include the random changepoint τ_i in both AD and death survival hazard functions, because accelerated decline in cognition is not only observed prior to AD diagnosis but also begins about 3 to 6 years prior to death (Wilson et al., 2003).

In the joint model we propose, all the subjects are considered to have the risk of developing the events of interest and will all experience a changepoint if they are followed long enough. That is we at this time do not consider cure rate model which assumes a null risk of developing event for a portion of the subjects. Bringing the cure fraction will increase the complexity

of the joint model by allowing some of the subjects have a linear trend and uncertainty of the existence of a random changepoint. We may consider it in future as an extension.

4.3.2 Bayesian Inference

Because the joint likelihood is quite complicated with a large number of parameters, we apply a hierarchical joint modeling framework for Bayesian inference to directly model the random changepoint and risk to AD.

We first start with the broken-stick model (1), which is the simplest one among the five. Note that the discontinuity at the breakpoint causes problems in asymptotic theory for frequentist approach, and Bayesian method of inference could avoid such unsatisfactory performance of asymptotics. For Bacon-Watts model (2), we use $trn = tanh$, i.e. the hyperbolic tangent, which makes the computation much easier. For Griffiths-Miller model (3), though the γ value could be predetermined as small as 0.1, which was suggested in Segalas et al., 2020's paper when applying on cognitive data, we here like to assign a uniform prior for it. In polynomial model (5), g is defined as a cubic polynomial with $g(x) = a_3x^3 + a_2x^2 + a_1x + a_0$. To solve the preceding linear system of four ordinal differential equations, we follow Yang and Gao, 2013's work and reduce $g(.)$ to a quadratic polynomial with the following coefficients: $a_2 = \frac{\tau_i - \beta_{1i}}{2\gamma}$, $a_1 = \beta_{1i} - \frac{\beta_{2i} - \beta_{1i}}{\gamma}\tau_i$, $a_0 = \beta_{0i} + \frac{\tau_i - \beta_{1i}}{2\gamma}(\beta_{2i})^2$. We will present comparisons of these models with both simulated and real data.

We assign $\pi(\theta)$ and $\pi(b)$ as the prior distributions for the unknown fixed effect parameters θ and random effects b. The joint prior can be specified as the product of the priors, which are independent of each other. The joint posterior probability distribution $f(\theta, b|Y, T)$ is proportional to $f(Y, T|\theta, b)\pi(\theta)\pi(b)$ with:

$Y_{ij}|\theta, b \sim N(f(t_i, \tau_i, \theta, b_i), {\sigma_y}^2)$

$\beta = (\beta_0, \beta_1, \beta_2) \sim N(\mu_\beta^\top, \sigma_\beta^2 I)$

$b_i = (b_{0i}, b_{1i}, b_{2i}) \sim N(0, \Sigma_b)$

$\tau_i \sim N(\mu_\tau, \sigma_{\tau_i}^2) I(min, max)$

$\beta_\tau = (\beta_{\tau_0}, \beta_{\tau_1}) \sim N(\mu_{\beta_\tau}^\top, \sigma_{\beta_\tau}^2 I)$

$\gamma_\tau \sim U(0, Max_\gamma)$

$\alpha = (\alpha_1, \alpha_2) \sim N(\mu_\alpha^\top, \sigma_\alpha^2 I)$

$\zeta \sim N(\mu_\zeta, \sigma_\zeta^2)$

$r(r_1, r_2, r_3) \sim N(\mu_r^\top, \sigma_r^2 I).$

The above specification can be modified regarding the selected changepoint formulation. To ensure the estimates are heavily relying on the observed data, we specify priors based on weakly informative distributions that are commonly implemented in the literature. For parameters of fixed effects (e.g. βs, β_τs, αs, ζ, and rs), we take normal distribution with zero mean and large variance (1000); for error terms, we use $Gamma^{-1}(0.01, 0.01)$. For variance-covariance matrix Σ_b of random effects, we take a inverse-Wishart distribution $\Sigma_b \sim Wishart^{-1}(3, R)$, where $R = diag(0.01, 3)$. In the cause-specific hazard function for a competing-risk setting, we have shape (a_k) and scale (λ_k) parameters for event k and they both follow a $Gamma^{-1}(0.01, 0.01)$ distribution.

Changepoint τ_i and transition parameter γ are critical in modeling the shape of the trend. An unbounded τ_i or γ may lead to a computational breakdown in the Bayesian MCMC process. We see as $\tau \to \infty$, $f(t_i, \tau_i, ; \theta, b)$ approaches a straight line. Too wide of the transition interval is meaningless. Careful consideration is required to choose a reasonable set of prior values for τ_i and γ. We assign a truncated normal distribution for the random changepoint with reasonable min/max values and bounded uniform distribution (0, max) for the transition parameter $\gamma_\tau > 0$.

The convergence of the MCMC samples of the parameters can be monitored by displaying trace plots and autocorrelations. Samples after excluding the initial burn-in will be diagnosed by standard methods, e.g. Gelman-Rubin test.

4.3.3 Simulation

Simulation studies allow the comparison of the performances of our proposed models. We simulate the longitudinal trajectory of a marker, according to the bent-cable model proposed by Khan and Kar, 2017. This model includes a quadratic bend in the change in slope to provide sufficient flexibility for the transition zone. The data generated from the bent-cable model could represent different types of changepoint data and parameters are easily interpretable, making it more realistic in practice. To make the model simple, we only consider the time effects (no other covariates) in the longitudinal model that $Y(t) = \beta_0 + b_0 + (\beta_1 + b_1)t + f(\beta_2, b_2, \tau, \gamma, t) + \epsilon_Y$, $\epsilon_Y \sim N(0, \sigma_Y^2)$, where the changepoint τ takes a normal distribution $N(\mu_\tau, \sigma_\tau^2)$ and the mean of τ is $\mu_\tau = \beta_{\tau 0} + \beta_{\tau 1} x$. Covariate x is a binary variable related to changepoint. Considering the complexity of the joint models with changepoint, long simulation time and our focus of choosing the best random changepoint function, here we only take a single event survival model without covariate. The association between longitudinal marker and time-to-event is depending on the changepoint τ. Random effects from longitudinal submodels are not included due to different meanings of the effects regarding different transition formulations in the random changepoint models. The simple linking structure is supported by previous proposed joint models by Jacqmin-Gadda et al., 2006 and Yu and Ghosh, 2010. Thus, the survival function is $S(t) = exp(-ht^a)$ with $h = h_0 exp(\alpha_0 + \zeta\tau)$.

The true parameter values are chosen based on the application on the real data (MAP) to allow reasonable generalization. Specifically for each subject, we use his baseline age

from MAP subtracted by 60 as the time at the first visit. The longitudinal measures are annually observed until the event or with maximum 15 years and the measurement times are assumed identical for all subjects. In all the scenarios, we take total subjects n = 400 with $t_i = 1, 2, ..., 15(max)$ for $i = 1, 2, ..., n$. For each simulation, 500 data sets are generated. We use 2000 MCMC iterations after burn-in to approximate posterior density. Posterior summaries are averaged over the 500 replicates for each parameter, and the means and coverage probabilities (proportion of such intervals out of 500 that capture the truth) is calculated.

We present simulation results for three scenarios by varying the transition parameter γ, linking parameter r and slope difference before/after changepoint.

4.3.3.1 Scenario 1

Since data may exhibit an either more or less gradual change in rate, it is important for the joint model to be flexible to predict the width of the transition. We generate data from a bent-cable model (quadratic bend) with (a) $\gamma = 2$ (4 years for the transition zone), (b) $\gamma = 0.5$ (1 year, close to a broken-stick). Note that, to reduce the model complexity, based on previous work of Yu and Ghosh, 2010 and Jacqmin-Gadda et al., 2006, we drop the random slope term b_1 before the changepoint because its estimation is close to 0 and 95% confidence interval covers 0. $\beta_0 = 1.0$ and slopes parameters are β_1 = -0.1 and β_2 = -0.5. The random effect parameters for b_i are: $\sigma_0^2 = 0.05$, $\sigma_2^2 = 0.005$, $Cov_{02} = 0.004$. The residual of longitudinal response Y and changepoint τ are 0.04 and 9, respectively. The mean of τ is $\beta_{\tau_0} + \beta_{\tau_1} x$ with $\beta_{\tau 0} = 88$ and $\beta_{\tau_1} = 2.0$. The binary covariate x follows $Binary(0.2)$. We assume Weibull distribution for survival function with a censoring rate of 0.40. We further assume the censoring mechanism is independent of both event risk and longitudinal profile. The linking parameter ζ in the survival function is set as -0.6 in this scenario. We then fit

the 5 joint model with different random changepoint functions to each of the simulated data sets. Tough different model has different parameterizations, our primary interest are the parameters for the changepoint (β_{τ_0} and β_{τ_1}), transition parameter γ, slopes of the 2 linear phases (β_1 and β_2), and the risk of event (α and ζ).

Numerical results are summarized in Table XII. Since the direct interpretation from Bacon-Watts and Griffith-Miller model for βs and γ are not available, the comparison for these parameters are limited for these two models. Also note that, for polynomial model, we use $\tau + \gamma/2$ as a new τ in the comparison, which is the actual changepoint based on the model defined above. We take $\beta_2 - \beta_1$ in the polynomial model is for the slope difference to compare with the β_2 in the generating model. For Scenario 1a (4 year-transition), the bent-cable model and polynomial regression perform pretty well with respect to bias and coverage rate. For other models, the posterior means are all close to the true parameter values, while coverage rates are not as good as bent-cable, especially for the variance of the main outcome. This is not unexpected as the bent-cable model is the generating model. In Scenario 1b, where the transition is designed to be within a year reflecting a very short transition period, we see a great improvement in the broken-stick model. In terms of bias and coverage probability, bent-cable and polynomial models provide similar estimates as those obtained using the broken-stick model. The results from scenario 1 suggest that the proposed bent-cable and polynomial regression joint models have the flexibility to describe the changepoint data that exhibit either an approximately abrupt or gradual transition between the two lines.

4.3.3.2 Scenario 2

In scenario 2, we assume a relatively weaker effect of the changepoint time on the risk of event by setting the linking parameter $\zeta = -0.2$ in the generated data. Other parameters

Table XII: SIMULATION STUDY RESULT FOR JOINT MODEL WITH DIFFERENT TRANSITION FUNCTIONS (SCENARIO 1A AND 1B)

			Broken-stick		Bacon-Watts		Griffith-Miller		Bent-cable		Polynomial	
Scenario	Parameter	True	Mean	CR%e	Mean	CR%	Mean	CR%	Mean	CR%	Mean	CR%
Scenario 1a	Longitudinal											
	β_0	1.0	1.40	(0.02)	-	(-)	-	(-)	1.00	(0.94)	1.00	(0.93)
	β_1	-0.1	-0.11	(0.02)	-	(-)	-	(-)	-0.10	(0.93)	-0.10	(0.94)
	β_2	-0.5	-0.47	(0.00)	-	(-)	-	(-)	-0.50	(0.90)	-0.50	(0.89)
	γ	2.0	-	(-)	-	(-)	-	(-)	2.00	(0.93)	2.00	(0.93)
	σ_ϵ	0.2	0.20	(0.71)	0.21	(0.01)	0.20	(0.37)	0.20	(0.95)	0.20	(0.95)
	Changepoint											
	β_{τ_0}	88.0	87.87	(0.87)	88.04	(0.94)	88.11	(0.88)	87.99	(0.95)	87.99	(0.95)
	β_{τ_1}	2.0	1.95	(0.97)	1.83	(0.95)	1.85	(0.95)	1.99	(0.97)	1.99	(0.97)
	σ_τ	3.0	2.95	(0.91)	2.81	(0.59)	2.85	(0.71)	3.00	(0.92)	3.00	(0.93)
	Survival											
	ζ	-0.6	-0.61	(0.92)	-0.59	(0.93)	-0.58	(0.89)	-0.60	(0.95)	-0.60	(0.96)
Scenario 1b	longitudinal											
	β_0	1.0	1.03	(0.93)	-	(-)	-	(-)	1.00	(0.95)	1.00	(0.96)
	β_1	-0.1	-0.10	(0.93)	-	(-)	-	(-)	-0.10	(0.96)	-0.10	(0.95)
	β_2	-0.5	-0.50	(0.88)	-	(-)	-	(-)	-0.50	(0.91)	-0.50	(0.90)
	γ	0.5	-	(-)	-	(-)	-	(-)	-0.50	(0.93)	0.48	(0.96)
	σ_ϵ	0.2	0.20	(0.94)	0.21	(0.09)	0.21	(0.10)	0.20	(0.94)	0.20	(0.94)
	Changepoint											
	β_{τ_0}	88	87.99	(0.95)	88.08	(0.92)	88.09	(0.90)	87.99	(0.96)	87.99	(0.94)
	β_{τ_1}	2	1.99	(0.96)	1.85	(0.95)	1.86	(0.95)	1.99	(0.96)	1.99	(0.96)
	σ_τ	3	3.00	(0.93)	2.84	(0.69)	2.85	(0.72)	3.00	(0.93)	3.00	(0.92)
	Survival											
	ζ	-0.6	-0.60	(0.95)	-0.60	(0.93)	-0.59	(0.93)	-0.60	(0.95)	-0.60	(0.95)

are the same as in scenario 1a. We compare the 5 models' fit and the results are shown in Table XIII top. We notice all the 5 models can provide accurate estimates with respect to the association parameters ζ. Compared to other models, bent-cable and polynomial models perform better in the case of other parameters.

4.3.3.3 Scenario 3

To assess the model sensitivity to the slope difference before and after the changepoint, we consider scenario 3 to generate data for which $\beta_1 = -0.1$ and $\beta_2 = -0.1$, reflecting an undramatic change in rate of decline. Other parameters are the same as in scenario 1a. In Table XIII bottom, we see the performance of the bent-cable model in comparison with the true values is quite satisfactory. The result from the polynomial model is slightly

Table XIII: SIMULATION STUDY RESULT FOR JOINT MODEL WITH DIFFERENT TRANSITION FUNCTIONS (SCENARIO 2 AND 3)

			Broken-stick		Bacon-Watts		Griffith-Miller		Bent-cable		Polynomial	
Scenario	Parameter	True	Mean	CR%	Mean	CR%	Mean	CR%	Mean	CR%	Mean	CR%
Scenario 2	longitudinal											
	β_0	1.0	1.40	(0.02)	-	(-)	-	(-)	1.00	(0.95)	1.00	(0.95)
	β_1	-0.1	-0.11	(0.02)	-	(-)	-	(-)	-0.10	(0.95)	-0.10	(0.95)
	β_2	-0.5	-0.47	(0.00)	-	(-)	-	(-)	-0.50	(0.90)	-0.50	(0.88)
	γ	2.0	-	(-)	-	(-)	-	(-)	2.00	(0.96)	2.00	(0.94)
	σ_ϵ	0.2	0.20	(0.70)	0.21	(0.02)	0.20	(0.34)	0.20	(0.94)	0.20	(0.95)
	Changepoint											
	β_{τ_0}	88.0	87.86	(0.86)	88.03	(0.94)	88.11	(0.89)	87.99	(0.95)	87.99	(0.94)
	β_{τ_1}	-2.0	1.95	(0.97)	1.79	(0.93)	1.81	(0.94)	1.99	(0.97)	1.99	(0.97)
	σ_τ	3	2.94	(0.90)	2.78	(0.49)	2.82	(0.63)	3.00	(0.94)	3.00	(0.94)
	Survival											
	ζ	-0.2	-0.20	(0.96)	-0.21	(0.92)	-0.21	(0.95)	-0.20	(0.96)	-0.20	(0.96)
Scenario 3	longitudinal											
	β_0	1.0	1.08	(0.86)	-	(-)	-	(-)	1.00	(0.97)	1.00	(0.95)
	β_1	-0.1	-0.10	(0.86)	-	(-)	-	(-)	-0.10	(0.96)	-0.10	(0.94)
	β_2	-0.1	-0.10	(0.78)	-	(-)	-	(-)	-0.10	(0.88)	-0.10	(0.86)
	γ	2.0	-	(-)	-	(-)	-	(-)	1.88	(0.93)	1.89	(0.91)
	σ_ϵ	0.2	0.20	(0.94)	0.21	(0.14)	0.21	(0.13)	0.20	(0.95)	0.20	(0.95)
	Changepoint											
	β_{τ_0}	88.0	87.88	(0.89)	84.73	(0.00)	84.5	(0.00)	87.99	(0.92)	87.99	(0.90)
	β_{τ_1}	-2.0	1.99	(0.96)	1.07	(0.10)	1.04	(0.08)	1.99	(0.96)	2.00	(0.96)
	σ_τ	3	3.00	(0.95)	1.73	(0.00)	1.69	(0.00)	3.01	(0.93)	3.01	(0.93)
	Survival											
	ζ	-0.6	-0.61	(0.89)	-0.72	(0.64)	-0.73	(0.62)	-0.62	(0.89)	-0.63	(0.86)

off compared to the bent-cable but still acceptable. The performance of the Bacon-Watts and Griffith-Miller models are not good with respect to both bias and coverage for the changepoint parameters, suggesting their inadequacy in detecting the changepoint when change in slope is not steep.

Figure 8 a-d are the boxplots of simulated bias and standard error for each estimator. Among all the models, bent-cable leads to the best with respect to both coverage and smaller bias. In most cases the average of the estimates are almost equal to the true values. The performance of the broken-stick model is not as good as bent-cable for scenario 1a, 2 and 3, suggesting its inadequacy in characterizing gradual transition over time. However, we see much improved performance of this piecewise linear model, nearly as well as the bent-cable model, when data exhibit a quick transition trend (scenario 1b). The Bacon-Watts

and Griffith-Miller models appear that they cannot handle the simulated data adequately, especially when the change in slope is not large. Polynomial model's performance is also quite satisfactory for all the scenarios but the convergence is noticeably slow due to the model complexity. Regarding the model's performance and estimation efficiency, we adopt the bent-cable model for our further analysis.

4.4 Extension: Joint Model with Bivariate Random Changepoints for Longitudinal Outcomes

Thus far, we have only examined the joint model with changepoint for one response variable. Clinical and health research shows that many biological changes begin to develop before the appearance of clinical symptoms. There have been a few studies on joint modeling the bivariate longitudinal outcomes with different random changepoints simultaneously. In this session, we propose a bivariate random changepoint model with a focus on the correlations between the two changepoints. The association of the longitudinal and risk of event through the changepoints is also interested. Above simulation study suggests a good performance of bent-cable function in estimating the parameters for random changepoint in proposed joint model. Motivated by data from MAP study, we develop a joint model for bivariate longitudinal outcomes using bent-cable describing smooth transition and further extended with competing risk process.

4.4.1 Framework of the Joint Model

Let Y_{1ij} and Y_{2ij} be the two longitudinal responses of subject i at t_j; $\beta_{10}, \beta_{11}, \beta_{12}, b_{10}, b_{11}$ and b_{12} be the fixed and random effects for Y_1 with $\beta_{20}, \beta_{21}, \beta_{22}, b_{20}, b_{21}$ and b_{22} for Y_2; τ_{1i} and τ_{2i} be the times at the changepoint of longitudinal response Y_1 and Y_2 of subject i,

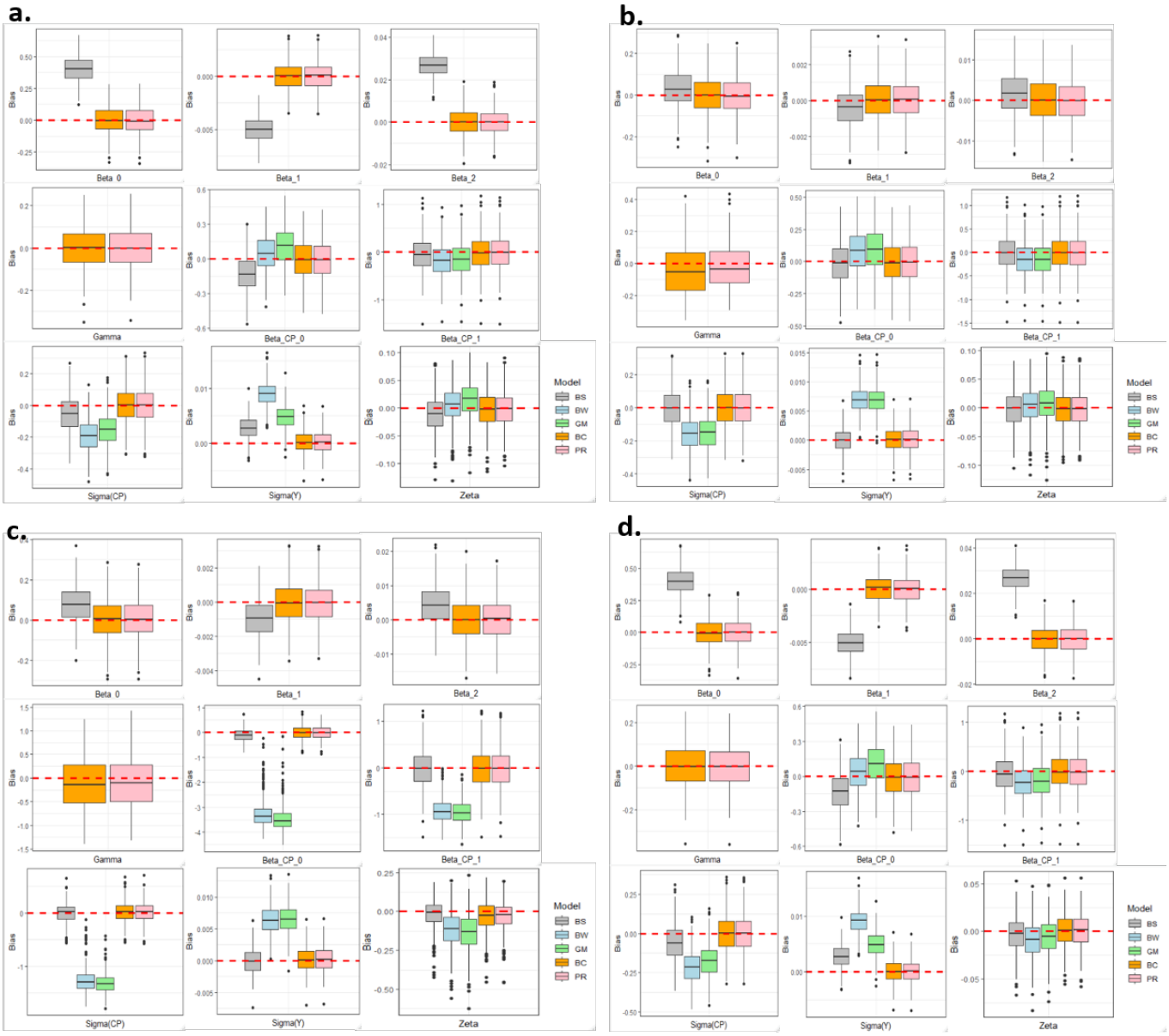

Figure 8: Boxplots of the estimated biases for joint models with different transition functions in simulation study
a.scenario 1a b.scenario 1b c.scenario 2 d.scenario 3

respectively; γ_1 and γ_2 be the transition parameters; ϵ_{1ij} and ϵ_{2ij} be the residual errors. Taking bent-cable function for its flexibility in characterizing a changepoint trajectory, we build the joint model of bivariate longitudinal outcomes as:

$$\begin{cases} Y_{1ij} = (\beta_{100} + \beta_{101}x_i + b_{10i}) + (\beta_{11} + b_{11i})t_{1ij} + \\ (\beta_{12} + b_{12i})\left(\dfrac{(t_{1ij} - \tau_{1i} + \gamma_1)^2}{4\gamma_1} I[\tau_{1i} - \gamma_1 < t_{1ij} \leq \tau_{1i} + \gamma_1] + (t_{1ij} - \tau_{1i}) I[t_{1ij} > \tau_{1i} + \gamma_1]\right) + \epsilon_{1ij} \\ Y_{2ij} = (\beta_{200} + \beta_{201}x_i + b_{20i}) + (\beta_{21} + b_{21i})t_{2ij} + \\ (\beta_{22} + b_{22i})\left(\dfrac{(t_{2ij} - \tau_{2i} + \gamma_2)^2}{4\gamma_2} I[\tau_{2i} - \gamma_2 < t_{2ij} \leq \tau_{2i} + \gamma_2] + (t_{2ij} - \tau_{2i}) I[t_{2ij} > \tau_{2i} + \gamma_2]\right) + \epsilon_{2ij}, \end{cases} \tag{4.15}$$

where, we allow the correlation between the two longitudinal markers random effects through following specific variance-covariance structure which is similar to (3.16 and 3.17):

$$b_i = \begin{pmatrix} b_{1i} \\ b_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} \Sigma_{b1} & \Sigma_{b12} \\ \Sigma_{b12}^{\top} & \Sigma_{b2} \end{pmatrix}\right). \tag{4.16}$$

The random changepoints τ_{1i} and τ_{2i} are modelled as jointly normally distributed with unknown mean expressed with covariate z_1 and z_2, respectively as:

$$\mu_{\tau_{1i}} = \beta_{\tau_{10}} + \beta_{\tau_{11}} z_{1i}; \quad \mu_{\tau_{2i}} = \beta_{\tau_{20}} + \beta_{\tau_{21}} z_{2i}, \tag{4.17}$$

and the correlation between the changepoints of τ_{1i} and τ_{2i} is:

$$\begin{pmatrix} \tau_{1i} \\ \tau_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\tau_{1i}} \\ \mu_{\tau_{2i}} \end{pmatrix}, D = \begin{pmatrix} \sigma^2_{\tau_1} & Cov_{\tau_{12}} \\ Cov_{\tau_{12}} & \sigma^2_{\tau_2} \end{pmatrix}\right). \tag{4.18}$$

To model the risk of each competing event, the proportional hazard model is constructed with full setting:

$$
\begin{aligned}
&h_k(t|h_k, \alpha_k, \zeta_k, r_k) = h_{k0}(t)exp[\alpha_k x_i + \zeta_{1k}\tau_{1i} + \zeta_{2k}\tau_{2i} + \\
&r_{11k}b_{10i} + r_{12k}b_{11i} + r_{13k}b_{12i} + r_{21k}b_{20i} + r_{22k}b_{21i} + r_{23k}b_{22i}], \\
&where\ k = 1, 2.
\end{aligned} \tag{4.19}
$$

The proposed model establish the association between longitudinal processes and the onset of clinical events. We need to be cautious when building up the linking between the longitudinal and survival processes using the shared parameters. Considering longitudinal responses are usually correlated and one predictor variable in the regression function may be linearly predicted from the other, adding both responses' changepoint and random effects as regressors in the proportional hazard function might lead to multicollinearity problem, which weakens the statistical power of the joint model. Also it is not practical with too many variables or random effects included, which will make the estimation process of the already complex joint model slow and run into a converge problem. The selection of the best model is most often processed by model comparison. Last, the interpretation of the results should be clinical meaningful and guided by scientific evidence as one longitudinal trajectory could be more specifically associated with one of the competing events than the other.

4.4.2 Bayesian Inference

Estimation of joint model with multiple random changepoints would be computationally challenging. Here we apply a hierarchical Bayesian inferential framework for the joint model with bivariate random changepoints. The prior distributions for most parameters are similar to those for the univariate changepoint joint model we proposed previously. They are fairly vague, minimally informative priors. Specifically in the bivariate changepoint framework,

for the correlation of the random effects in the longitudinal models, we chose a inverse-Wishart prior distribution $Wishart^{-1}(6, R)$, where $R = diag(0.001, 6)$; for the jointly modeled changepoints, we use variance $(\sigma^2_{\tau_1}, \sigma^2_{\tau_2})$ and correlation coefficient ρ to represent their variance-covariance matrix $\begin{pmatrix} \sigma^2_{\tau_1} & \rho\sigma_{\tau_1}\sigma_{\tau_2} \\ \rho\sigma_{\tau_2}\sigma_{\tau_1} & \sigma^2_{\tau_2} \end{pmatrix}$. The prior for ρ follows uniform distribution $U(-1, 1)$. ρ reflects the correlation between the changepoints of the two outcomes and it is mostly likely to be positive as the two longitudinal responses are under a similar tendency. The priors for the linking parameters $r(r_{11}, r_{12}, r_{13})$ each follows $N(0, 1000)$ for outcome Y_1 and $r(r_{21}, r_{22}, r_{23})$ follows $N(0, 1000)$ for outcome Y_2. The choice of shared random elements in each hazard function depends on the statistical feasibility and clinic meaningfulness. We assume the linking framework for bivariate longitudinal markers and two competing events is either relying on just one longitudinal process or achieved by assigning different longitudinal trajectories to different events.

We obtain the estimates using MCMC method implemented by WinBUGS. The convergence of the MCMC samples takes a noticeable longer time compared to the univariate random changepoint joint model.

4.4.3 Simulation

We assess the performance of the proposed method for bivariate random changepoint joint model in simulation studies. We would like to investigate how well the parameters are estimated, and particularly, we are interested in assessing the performance of estimators of the association parameters under different scenarios.

We generate data for N=400 subjects with two correlated longitudinal outcomes (Y_1 and Y_2) and two competing events ($k = 1$ and 2). The simulated data is based on the estimated parameters from fitting the joint model with bent-cable random changepoints to

the real data. Parameter values are chosen to allow reasonable generation. We assume the measurement times are annual with maximum follow-up of 15 years and visit time intervals are identical for all subjects. For the first longitudinal marker (Y_1), we have $\beta_{10} = 1$, $\beta_{11} = -0.1$, $\beta_{12} = -0.5$ and residual variance $\sigma^2_{\epsilon_1} = 0.04$, while for the second marker (Y_2) $\beta_{20} = 2$, $\beta_{21} = -0.2$, and $\beta_{22} = -0.6$, and $\sigma^2_{\epsilon_2} = 0.25$. Variance parameters for b_i is given as: $\Sigma_{b1} = \begin{pmatrix} 0.05 & 0.004 \\ 0.004 & 0.005 \end{pmatrix}, \Sigma_{b2} = \begin{pmatrix} 0.02 & -0.002 \\ -0.002 & 0.02 \end{pmatrix}, \Sigma_{b12} = \begin{pmatrix} 0.01 & -0.001 \\ 0.004 & 0.001 \end{pmatrix}$. The transition parameters γ_1 and γ_2 are set as 2 (4 years transition) and 1 year (2 years transition) for Y_1 and Y_2, respectively. For changepoints τ_1 and τ_2, we assume $\beta_{\tau_{10}} = 90$ and $\beta_{\tau_{11}} = 2$ for the first marker while $\beta_{\tau_{20}} = 86$ and $\beta_{\tau_{21}} = -2$ for the second, respectively. The binary covariates related to μ_{τ_1} and μ_{τ_2} are $z_1 \sim Binary(0.2)$ and $z_2 \sim Binary(0.3)$, respectively. We consider two scenarios for the correlation between the changepoints: the first (Scenario 1) corresponds to highly correlated change times with variance covariance matrix $\begin{pmatrix} 9 & 10 \\ 10 & 16 \end{pmatrix}$ where correlation coefficient $\rho = 0.83$ while the second one (Scenario 2) has $\begin{pmatrix} 9 & 4 \\ 4 & 16 \end{pmatrix}$ where correlation coefficient $\rho = 0.33$ for a weak correlation. For competing risks, we assume constant baseline cause-specific hazards using Weibull distribution, which yields, on average, 17% of the subjects with event 1 and 44% with event 2 and 39% are censored. We further assume censoring is non-informative. That is, the censoring mechanism is independent of both competing risks and longitudinal processes. We keep association parameters ζ_1 and ζ_2 for changepoints in the hazard functions, assuming random changepoint τ_1 is negatively associated with event 1 ($\zeta_1 = -0.5$), and τ_2 is negatively associated with event 2 ($\zeta_2 = -1.0$).

In addition, to assess the inference sensitivity to prior specifications for γ and τ, in Scenario 3 we change $\gamma \sim U(0, 5)$ to $U(0, 1)$ (3a) or $U(0, 10)$ (3b), representing a relatively narrow or wider interval for transition, respectively. We also consider to release the boundaries of the truncated normal distribution of $\tau_i \sim N(\mu_\tau, \sigma^2_\tau)$.

For each simulation, 500 data sets are generated. 2000 MCMC iterations after burn-in are used to approximate posterior density. Posterior summaries are averaged over the 500 replicates for each parameter, and the means and coverage probabilities (proportion of such intervals out of 500 that capture the truth) is calculated.

Table XIV, XV and XVI summarize the simulation results of the joint model with bivariate random changepoints for scenario 1,2 and 3a. Figure 9 a-c are the boxplots of simulated bias and standard error of the estimates. We focus on the following parameters: intercepts, slopes and variance of the two longitudinal markers, correlation between the random effects, time of the changepoints, variances of changepoints, and correlations between changepoints, linking parameters between longitudinal and survival processes. As we can see, the overall estimation quality is good with satisfying posterior means and interval coverage under the assumption of high ($\rho = 0.83$, Scenario 1) or low ($\rho = 0.33$, Scenario 2) correlation between the two changepoints. We also find that when γ has a wider prior distribution (Scenario 3b), the estimation procedure performs well (Appendix D). However, a narrow γ prior distribution ($\gamma \sim U(0, 1)$) leads to a poor coverage and larger bias for the parameters inferencing about the shape of the longitudinal change (βs and γs), while other parameters estimation remain acceptable (Table XVI). These results show that the proposed Bayesian model is roughly robust when the true value of γ falls within the interval of uniform prior, that is, the true $\gamma <$ upper bound. Note that since γ is positive, a reasonable setting of the lower bound of is 0. Moreover, further releasing the boundaries of the τ has no effect on the estimation result. Though boundary seems not necessary for τ in the simulation study where τ_i is simulated with an ideal normal distribution, we still suggest resealable boundary values for τ to improve the estimation efficiency in real data.

Table XIV: SIMULATION STUDY RESULT FOR JOINT MODEL WITH BENT-CABLE FUNCTION (SCENARIO 1)

Parameter	True	Post mean	SD	CR%
Longitudinal				
Outcome1				
β_{10}	1.0	1.00	0.09	(0.93)
β_{11}	-0.1	-0.10	0.00	(0.93)
β_{12}	-0.5	-0.50	0.01	(0.92)
γ_1	2.0	2.01	0.10	(0.95)
$\sigma_{\epsilon 1}$	0.2	0.20	0.002	(0.95)
Outcome2				
β_{20}	2	1.98	0.30	(0.92)
β_{21}	-0.2	-0.20	0.00	(0.92)
β_{22}	-0.6	-0.60	0.00	(0.93)
γ_2	1.0	0.91	0.27	(0.92)
$\sigma_{\epsilon 2}$	0.5	0.50	0.005	(0.90)
Cov_{10_20}	0.010	0.0101	0.0032	(0.90)
Cov_{10_21}	-0.001	-0.0009	0.0006	(0.93)
Cov_{11_20}	0.004	0.0043	0.003	(0.88)
Cov_{11_21}	0.001	0.0009	0.0003	(0.90)
Changepoint1				
$\beta_{\tau_1 0}$	90.0	90.00	0.17	(0.94)
$\beta_{\tau_1 1}$	2.0	2.02	0.24	(0.95)
σ_{τ_1}	3	3.02	0.12	(0.96)
Changepoint2				
$\beta_{\tau_2 0}$	86.0	85.99	0.23	(0.96)
$\beta_{\tau_2 1}$	-2.0	-1.99	0.23	(0.96)
σ_{τ_2}	4	4.02	0.15	(0.97)
ρ	0.83	0.83	0.02	(0.95)
Survival				
Event1				
ζ_1	-0.5	-0.52	0.07	(0.92)
Event2				
ζ_2	-1.0	-1.00	0.05	(0.90)

Table XV: SIMULATION STUDY RESULT FOR JOINT MODEL WITH BENT-CABLE FUNCTION (SCENARIO 2)

Parameter	True	Post mean	SD	CR%
Longitudinal				
Outcome1				
β_{10}	1.0	1.01	0.09	(0.93)
β_{11}	-0.1	-0.10	0.00	(0.94)
β_{12}	-0.5	-0.50	0.01	(0.92)
γ_1	2.0	2.01	0.10	(0.94)
$\sigma_{\epsilon 1}$	0.2	0.20	0.00	(0.94)
Outcome2				
β_{20}	2	1.99	0.30	(0.91)
β_{21}	-0.2	-0.20	0.00	(0.91)
β_{22}	-0.6	-0.60	0.01	(0.92)
γ_2	1.0	0.89	0.26	(0.94)
$\sigma_{\epsilon 2}$	0.5	0.50	0.00	(0.91)
Cov_{10_20}	0.010	0.0102	0.0033	(0.91)
Cov_{10_21}	-0.001	-0.0009	0.0006	(0.94)
Cov_{11_20}	0.004	0.004	0.0026	(0.87)
Cov_{11_21}	0.001	0.0008	0.0003	(0.88)
Changepoint1				
$\beta_{\tau_1 0}$	90.0	90.00	0.18	(0.94)
$\beta_{\tau_1 1}$	2.0	2.02	0.39	(0.94)
σ_{τ_1}	3	3.01	0.13	(0.94)
Changepoint2				
$\beta_{\tau_2 0}$	86.0	85.98	0.27	(0.97)
$\beta_{\tau_2 1}$	-2.0	-1.98	0.37	(0.96)
σ_{τ_2}	4	4.02	0.15	(0.95)
ρ	0.33	0.33	0.05	(0.95)
Survival				
Event1				
ζ_1	-0.5	-0.51	0.05	(0.94)
Event2				
ζ_2	-1.0	-1.01	0.06	(0.90)

Table XVI: SIMULATION STUDY RESULT FOR JOINT MODEL WITH BENT-CABLE FUNCTION (SCENARIO 3A)

Parameter	True	Post mean	SD	CR%
Longitudinal				
Outcome 1				
β_{10}	1.0	1.20	0.09	(0.32)
β_{11}	-0.1	-0.10	0.00	(0.27)
β_{12}	-0.5	-0.48	0.01	(0.02)
γ_1	2.0	0.98	0.00	(0)
$\sigma^2_{\epsilon 1}$	0.2	0.20	0.002	(0.92)
Outcome 2				
β_{20}	2	2.03	0.29	(0.91)
β_{21}	-0.2	-0.20	0.00	(0.91)
β_{22}	-0.6	-0.60	0.00	(0.91)
γ_2	1.0	0.69	0.14	(0)
$\sigma_{\epsilon 2}$	0.5	0.50	0.005	(0.91)
Random effects				
Cov_{10_20}	0.010	0.0103	0.0032	(0.91)
Cov_{10_21}	-0.001	-0.0008	0.0007	(0.95)
Cov_{11_20}	0.004	0.0044	0.0029	(0.80)
Cov_{11_21}	0.001	0.0009	0.0003	(0.87)
Changepoint 1				
$\beta_{\tau_1 1}$	2.0	1.99	0.24	(0.81)
σ_{τ_1}	3	2.96	0.12	(0.94)
Changepoint 2				
$\beta_{\tau_2 0}$	86.0	85.99	0.23	(0.95)
$\beta_{\tau_2 1}$	-2.0	-1.99	0.23	(0.96)
σ_{τ_2}	4	4.01	0.15	(0.95)
ρ	0.83	0.83	0.02	(0.96)
Survival				
Event 1				
ζ_1	-0.5	-0.53	0.07	(0.89)
Event 2				
ζ_2	-1.0	-1.01	0.06	(0.88)

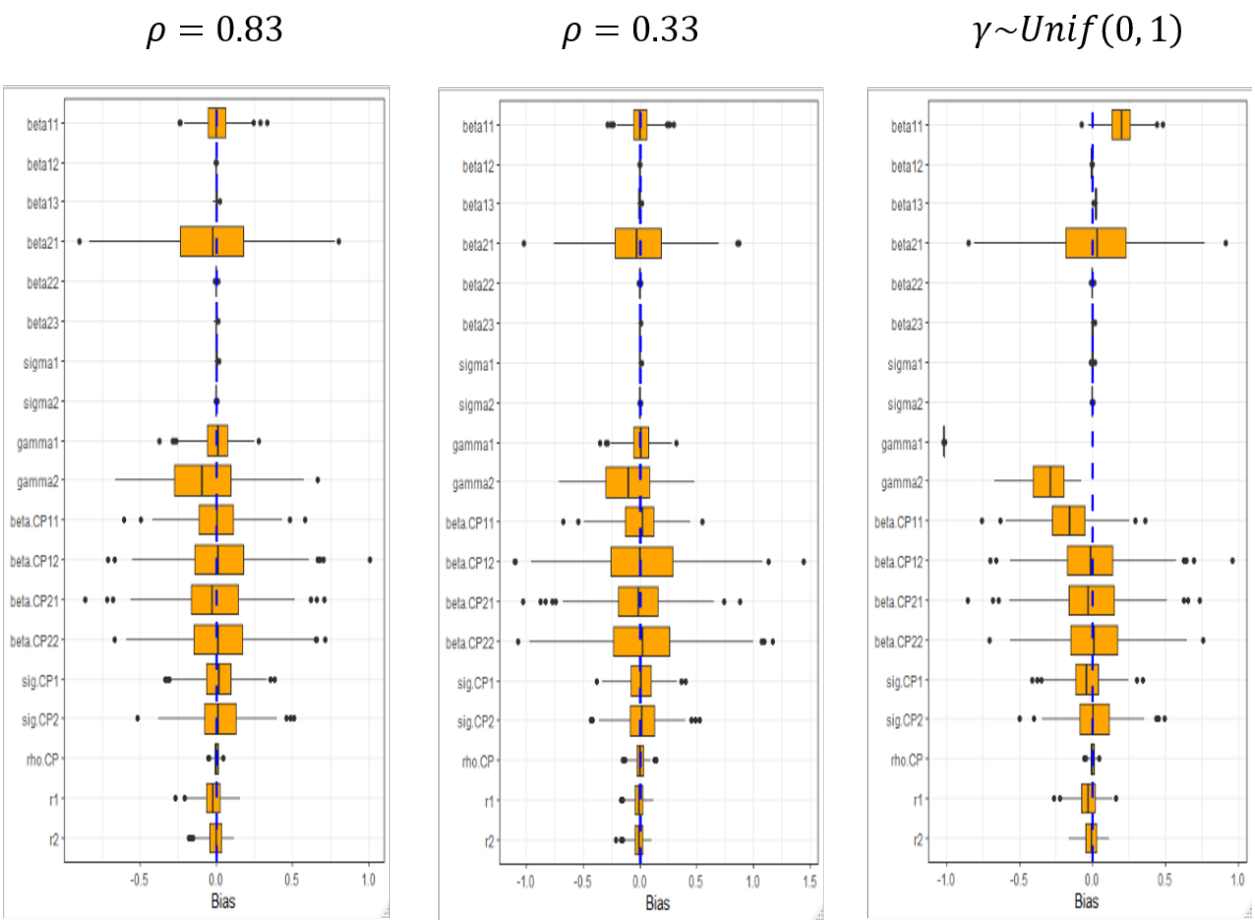

Figure 9: Boxplots of estimated biases for joint model with bivariate random changepoints in simulation study
a.scenario 1 b.scenario 2 c.scenario 3a

4.5 Discussion

In this chapter, to describe a changepoint trajectory of the longitudinal marker, we investigated 5 models with random changepoint coupled with survival process with competing risks. To our knowledge there are few articles comparing the various formulations of the random changepoint models using Bayesian approach, and this is the first time to combine the random changepoint model for a smooth transition with time-to-event data. Also, the changepoint is subject-specific with mean related to some covariates.

We carried out a simulation study to examine the robustness and estimating accuracy of each changepoint joint model. The broken-stick model has limit use due to its lack of realism. Bacon-Watts and Griffith-Miller model showed larger bias and poorer posterior interval coverage. Moreover, interpretation of βs and γ of these two models is not straightforward, as these parameters are linked with the shape of the transition. The polynomial model takes considerably longer computing time, making it relatively harder to implement compared to the bent-cable model.

To account for the variability in all the phases of the longitudinal trajectories and their association with time-to-event data, we extended the joint models with bivariate random changepoints. We assumed the correlation between the random changepoints as well as the subject-specific random intercept and slope. The simulation study confirmed a good performance of the proposed joint model. In addition, the proposed methodology is applicable to studies that is interested in investigating the order of the changepoints of multiple markers. The model would allow the comparison of the time of changepoint in slope between two longitudinal markers and provide a useful framework to assess their temporal order.

Our methodology provides a flexible approach to model changepoint trajectories, and some caution is required. First, γ plays an important role in modeling the shape and width

of the interval, it is necessary to have its prior distribution to be adequate to describe the transition phase. In practice, a time-series plot is useful to roughly determine the shape and timeline of the slope change. Knowledge based on previous studies could be necessary on choosing the priors (informative priors) and realistic initial values for τ and γ to overcome the computational difficulties for parameter estimation.

Second, most random changepoint models in previous studies used timescale interpreted as a delay to event onset, e.g. years before AD dementia diagnosis. It is good as long as when only the longitudinal trajectories are investigated. However, when we include terms regarding survival submodels to avoid a selection bias, the joint model approach makes it necessary to use age or follow-up time from baseline as timescale because the time to event cannot be determined for censored cases in such design. With a prospective timescale, the joint model can also be used to make predictions of the event.

Third, the association structure between the longitudinal marker trajectories and event risks relies on the shared random variables. The choice of random elements in each hazard function is depending on the statistical feasibility and clinic meaningfulness. Assuming the related trajectories of the two longitudinal outcomes with changepoints correlated at some degree, it is not a good practice to have the random changepoint from both longitudinal models in the same survival function. We could apply some model-selection criterion to choose an appropriate linking framework for the joint model.

Last, the random changepoint methodology we introduced here is intended for data that exhibit only one transition period over time. For models with multiple changepoints, the estimation would be a computational challenge. For a smoother model, we may consider a smooth function, for example B-spline model, which is capable of capturing the nonlinear evolution but has difficulties when interpreting the results due to multiple basis functions. An example of this approach is the work by Mokhles et al., 2012.

5 APPLICATION ON MAP DATA

In the above section, we introduced the 5 random changepoint models with different frameworks describing the transition and discussed their properties. The simulation study results suggest the joint model with bent-cable transition function provides satisfactory performance in modeling different types of changepoint data. Now we demonstrate the application of these joint models in the MAP cohort. We first apply the methodology to assess the existence of a random changepoint in cognitive function and motor function trajectories by joint model with univariate longitudinal outcome. Then we model these two measures for both changepoints in a joint modeling framework coupled with health-event data. We restrict the data to participants having at least five cognitive function assessments (N = 717). The changepoint, τ_i, is the age of the center in the bent zone between the two linear phases measured in years, and is allowed to be different across subjects.

5.1 Joint Model with Changepoint in Cognitive Decline and Risk of AD in the Presence of Death

As the MAP data provide information for the deceased participants, adding the competing risk of death to the joint modeling framework is available. The model has two events: AD and AD free death. Compared to Chapter 3, in the section we take into account the changepoint in cognitive trajectory and the time to death, this approach will provide a better fit to the data and more accurate description of the association between longitudinal outcomes and progression of AD.

5.1.1 Model Specification

The joint model is formulated under Bayesian framework for statistical inference. Because joint models are applied to the whole population including events and censored data, we use age as the timescale. We first specify the longitudinal model for the trajectory of cognitive function. Gender and education are used as the covariates for cognitive decline. Subject-specific changepoint τ_i is the age at acceleration of the decline (center of the transition zone) and follows a truncated normal with range from 50 to 105 years. The mean μ_{τ_i} is assumed related to education and ApoE4 as $\mu_{\tau_i} = \beta_{\tau_0} + \beta_{\tau_1} edu_i + \beta_{\tau_2} ApoE4_i$. Note that, in order to reduce the model complexity, following the previous work of Yu and Ghosh, 2010 and Jacqmin-Gadda et al., 2006, we drop the random effect for phase 1 slope before the changepoint because its estimation is close to 0 and 95% confidence interval covers 0. In the final model for cognitive function, we have

$$Globcog_{ij} = (\beta_{00} + \beta_{01} gender_i + \beta_{02} edu_i + b_{0i}) + \beta_1 age_{ij} + (\beta_2 + b_{1i}) * q(age_{ij}; \tau_i, \gamma) + \epsilon_{ij},$$

where

$$q(age_{ij}; \tau_i, \gamma) = \left(\frac{(age_{ij} - \tau_i + \gamma)^2}{4\gamma} I[\tau_i - \gamma < age_{ij} \leq \tau_i + \gamma] + (age_{ij} - \tau_i) I[age_{ij} > \tau_i + \gamma] \right). \tag{5.1}$$

Time in the cause-specific hazard survival model is the age of diagnosis of AD or age at death, and we choose education and ApoE4 as covariates in the AD subsurvival function (h_{AD}) while gender and education are in AD-free death subsurvival function (h_D). Moreover, because of the time gap between evaluating visits, the AD onset time is considered as interval censored between the diagnosed visit and the visit before $(t_{i(last-1)}, t_{i(last)})$. The link between longitudinal and survival processes depends on the subject-level age at changepoint τ_i, random effects from cognitive trajectory b_i. For a better model convergence, we center

the age at changepoint by subtracting 90. The hazard function of the survival submodel is:

$$\begin{cases} h_{AD}(t_{ADi}) = h_{AD0}(t)exp[\alpha_{AD1}edu_i + \alpha_{AD2}ApoE4_i + \zeta_{AD}(\tau_i - 90) + r_{AD1}b_{0i} + r_{AD2}b_{1i}], \\ h_D(t_{Di}) = h_{D0}(t)exp[\alpha_{D1}gender_i + \alpha_{D2}edu_i + \zeta_D(\tau_i - 90) + r_{D1}b_{0i} + r_{D2}b_{1i}]. \end{cases} \tag{5.2}$$

We adopt Bayesian method of inference and weakly informative priors are used. Particularly, the prior for each parameter in $(\beta, \beta_\tau, \alpha, \zeta, r)$ has normal distribution with mean 0 and variance 1000. We assign inverse gamma (0.001,0.001) distribution for error term variances. For random effects variance-covariance matrix, we take a inverse-Wishart distribution $Wishart^{-1}(2, R)$, where $R = diag(0.001)$. Transition parameter γ follows a uniform distribution of (0, 5).

We construct Markov chains to approximate the posterior density. The converging behavior of the chain is monitored by trace and density plots. The length of the the burn-in and convergence of the Markov chain is determined by Gelman-Rubin diagnostic test (Appendix E).

5.1.2 Result

We summarize the posterior mean, standard deviation and 95% credible interval of the parameters in the joint model of cognitive decline with bent-cable function for a changepoint and AD incidence with competing risk of death in Table XVII. As it can be seen, both gender and education are highly correlated with baseline cognitive function score. The posterior mean of the slope before changepoint is almost flat ($\beta_1 = 0.006$) and the slope increases dramatically after the changepoint about 50 fold ($\beta_2 = 0.30$), suggesting the existence of an acceleration of the cognitive decline in the cognition trajectory. The transition parameter γ is 2.1, indicating a bent zone of approximately 4 years based on the definition in bent-

cable transition function. The average age at changepoint is $\overline{\tau_i} = 89.0$ for the sample of 717 subjects. Under the formulation, on average, the transition begins at age 86.9 (89.0-2.1) and ends at age 91.1 (89.0+2.1) with 89.0 in the center. Figure 10 displays the estimated subject-specific trajectories for 9 randomly selected subjects. We see a good individual fit with predicted trajectories matching the observed cognitive scores (blue circles) using bent-cable transition function in the joint model.

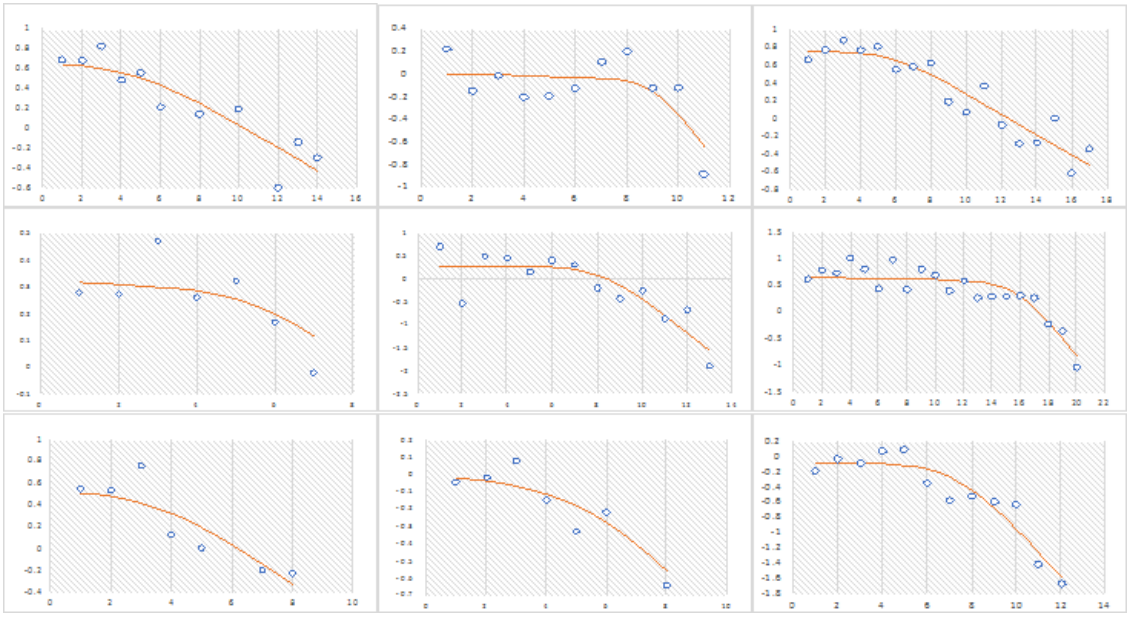

Figure 10: Individual observations and predicted cognition trajectories for randomly selected subjects
Blue circle: observed value; Red line: predicted value.

We assume the changepoint τ_i follows a normal distribution $N(\mu_{\tau_i}, \sigma^2_\tau)$, and the result shows μ_{τ_i} is related to ApoE4 ($\beta_{\tau_2} = -3.06$) but not education level. The age at the acceleration of the cognitive decline is about 3 years earlier for ApoE4 allele carriers compared to non-carriers. Figure 11 shows the estimated curves of the mean cognitive score given the averaged other covariate values for case of ApoE carriers and non-carriers, highlighting the differential evolution of the cognitive performance in the years prior to diagnosis according to the ApoE4 genotype. Before acceleration starts, the two cases have similar cognitive function

regardless of ApoE4. The changepoints are approximately at ages 86 (84-88 for transition zone) and 89 (87-91 for transition zone) for ApoE4 carriers and non-carriers, respectively. The trajectories are parallel by ApoE4 after the transition zone.

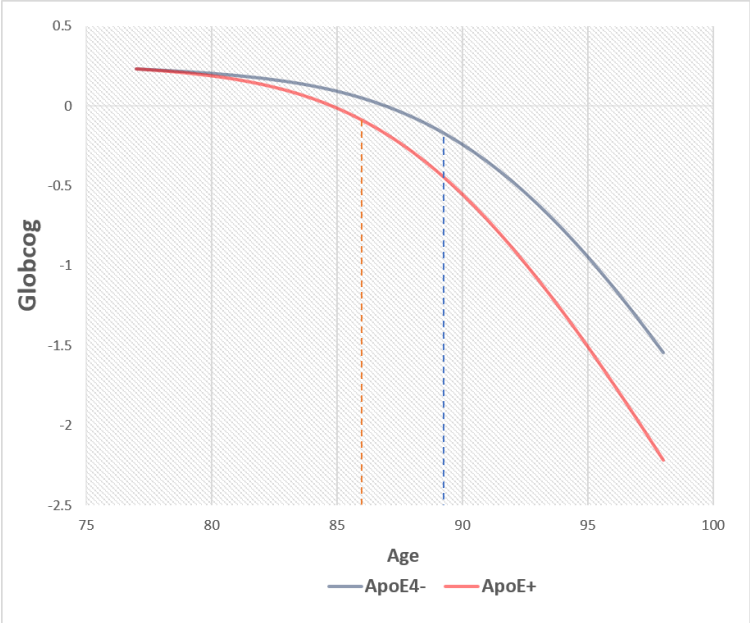

Figure 11: Expected cognitive score in the years before the diagnosis of AD by ApoE4 Red line: ApoE4+, acceleration starts at age of 84 ($\tau = 86$); Blue line: ApoE4-, acceleration starts at age of 87 ($\tau = 89$).

The survival process is modeled jointly with cognitive trajectory through a series of linking parameters, in which ζ_{AD} and ζ_{Death} are the parameters for the changepoint associated with risk of AD and death, respectively. Both ζs are negative ($\zeta_{AD} = -0.69$ and $\zeta_{AD} = -0.49$) with credible intervals not covering 0, indicating subjects with later age at changepoint has lower risk of AD or death comparing to the ones with earlier change in cognitive decline. Random intercept and slope terms are both significant in the AD survival submodel ($r_1 = -1.82$ and $r_2 = -5.06$). The coefficients are both negative, indicating lower cognition score at baseline or faster decline rate in the accelerated phase is associated with a higher risk of AD incidence. This is consistent with our result in Chapter 3 from the joint model with shared random effects and the results from other previous studies. The AD subsurvival model

Table XVII: PARAMETER ESTIMATES OF JOINT ANALYSIS OF COGNITIVE DECLINE WITH BENT-CABLE TRANSITION FUNCTION FOR A CHANGEPOINT AND AD INCIDENCE WITH COMPETING RISK OF DEATH

Parameter	Post mean	SD	95%CI
Longitudinal			
Intercept (β_{00})	-0.04	0.11	(-0.24, 0.22)
Gender (β_{01})	-0.09	0.03	(-0.15, -0.02)
Edu (β_{02})	0.06	0.01	(0.05, 0.07)
Age (β_1)	-0.006	0.001	(-0.008, -0.005)
Age (β_2)	-0.30	0.02	(-0.35, -0.27)
Transition (γ)	2.10	0.23	(1.64, 2.57)
Ran_int (σ_0^2)	0.12	0.01	(0.11, 0.14)
Ran_cov (Cov_{01})	0.007	0.006	(-0.004, 0.018)
Ran_age (σ_1^2)	0.04	0.00	(0.03, 0.05)
Error (σ_ϵ^2)	0.03	0.00	(0.03, 0.03)
Changepoint			
Intercept (β_{τ_0})	89.70	0.44	(88.87, 90.58)
Edu (β_{τ_1})	-0.20	0.57	(-1.31, 0.93)
ApoE4 (β_{τ_2})	-3.06	0.69	(-4.42, -1.71)
Error (σ_τ^2)	34.70	2.59	(29.89, 40.12)
Survival			
Event=AD			
Edu (α_1)	0.01	0.04	(-0.08, 0.07)
ApoE4 (α_2)	0.18	0.26	(-0.35, 0.68)
changepoint (ζ)	-0.69	0.03	(-0.74, -0.64)
Ran_int (γ_1)	-1.82	0.37	(-2.51, -1.07)
Ran_age (γ_2)	-5.06	0.64	(-6.34, -3.89)
Event=death			
Male (α_1)	0.71	0.25	(0.22, 1.22)
Edu (α_2)	-0.06	0.04	(-0.14, 0.01)
changepoint (ζ)	-0.49	0.03	(-0.54, -0.43)
Ran_int (γ_1)	-0.06	0.35	(-0.64, 0.76)
Ran_age (γ_2)	-10.56	1.04	(-12.73, -8.65)

shows ApoE4 genotype increases the risk of AD, but the effect is not significant. Since we have shown the risk of AD is highly correlated with the changepoint, which is associated with ApoE4, we like to plot the AD-free probability function by ApoE4 in Figure 12. Compared to non-carries, ApoE4 carriers have an earlier decline in survival rate, which starts around the beginning of the acceleration of the cognitive decline (age of 84). For competing risk of death, only random slope in the accelerating phase (Phase 2) is negatively associated with the risk of death ($r_2 = -10.56$). This is supported by recent study by Lv et al., 2019, where they found the association of acceleration in cognitive decline with mortality and this association was independent of initial cognitive function. We are not surprised to see males having higher risk of death ($\alpha_1 = 0.71$) as many research have shown that women have lower mortality rates compared to men. Our data also shows higher education level decreases the risk of death ($r_2 = -0.06$) and effect is marginally significant.

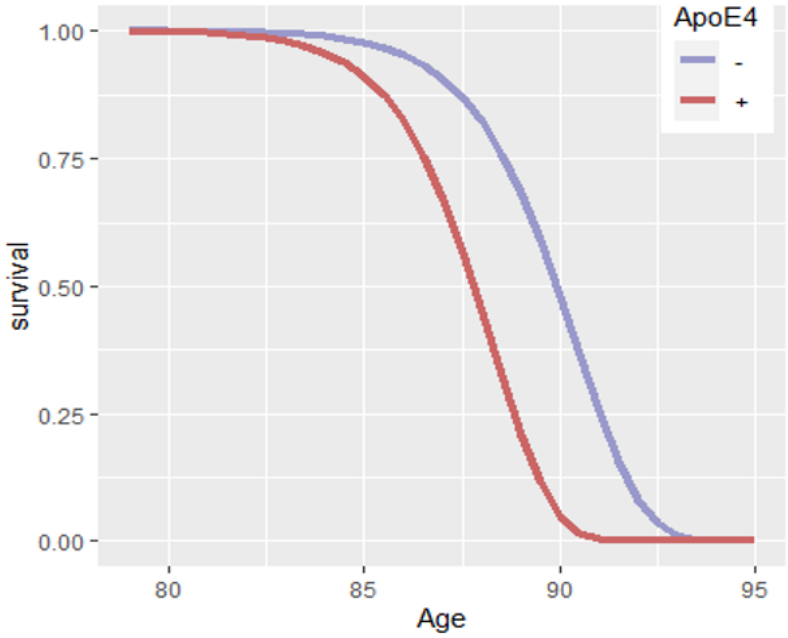

Figure 12: AD-free probability in years by ApoE4
Red line: ApoE4+; Blue line: ApoE4-

Table XVIII: PARAMETER ESTIMATES OF JOINT ANALYSIS OF COGNITIVE DECLINE WITH DIFFERENT TRANSITION FUNCTIONS FOR A CHANGEPOINT AND AD INCIDENCE WITH DEATH COMPETING RISK

	Broken-stick WAIC=-1424.7 $\overline{\tau}$=88.50(SD=4.80)		Bacon-Watts WAIC=-1439.2 $\overline{\tau}$=90.33(SD=4.30)		Griffith-Miller WAIC=-1489.0 $\overline{\tau}$=90.81(SD=4.43)		Bent-cable WAIC=-1492.2 $\overline{\tau}$=89.01(SD=4.74)		Polynomial WAIC=-1449.3 $\overline{\tau}$=89.09(SD=4.80)	
Parameter	mean	95%CI	mean	95%CI	mean	95%CI	mean	95%CI	mean	95%CI
$\beta_{1_{Age\_bf\_CP}}$	-0.008	(-0.010,-0.006)	-	(-)	-	(-)	-0.006	(-0.008,-0.005)	-0.006	(-0.008,-0.004)
$\beta_{2_{Age\_af\_CP}}$	-0.25	(-0.27,-0.23)	-	(-)	-	(-)	-0.30	(-0.35,-0.27)	-0.28	(-0.31,-0.25)
γ	-	(-)	0.29	(0.20,0.49)	2.13	(0.85,3.43)	2.10	(1.64,2.57)	4.10	(3.40, 4.82)
$\beta_{\tau_{ApoE4}}$	-2.83	(-4.19,-1.53)	-3.06	(-4.38,-1.70)	-2.57	(-3.89,-1.24)	-3.06	(-4.42,-1.71)	-3.29	(-4.68,-1.93)
ζ_{AD}	-0.64	(-0.71,-0.57)	-0.25	(-0.29,-0.21)	-0.73	(-0.78,-0.69)	-0.69	(-0.74,-0.64)	-0.42	(-0.46,-0.39)
ζ_{Death}	-0.39	(-0.46,-0.33)	-0.17	(-0.21,-0.13)	-0.40	(-0.45,-0.35)	-0.49	(-0.64,-0.43)	-0.27	(-0.31,-0.23)

Our model assumes that every subject develops AD eventually and consists of a changepoint in cognitive decline, and we notice that the estimated age at changepoint is not always observed before the end of the follow-up, especially for the censored. Out of 717 subjects, we have 132 diagnosed with AD, 200 deceased without AD and 385 censored. Among the 132 AD cases, the estimated changepoints by our proposed joint model are all (100%) taking place before the diagnosis of AD (last visit), in which 126 (95.5%) are during the follow-up and 6 are before their enrolled time. The changepoint is found to happen on average 4.02 (SD = 1.73) years before the diagnosis of AD. For 200 AD-free deceased cases, we have 160 (80.0%) cases with changepoint detected before death and we know death could be due to non-natural causes or from acute abdominal conditions that could be the explanation for part of the 20% with changepoint estimated after death. The averaged time between age at changepoint and death is 2.12 (SD = 3.21) years. For the 385 censored cases, we have 96 (24.9%) subjects with changepoint estimated to happen before their end points. These results are in agreement with previous research conclusion that acceleration in cognitive decline is several years prior to dementia and death.

We also run the joint models using other transition functions as well as the bent-cable. The important parameter estimates are summarized in Table XVIII. We see that the averaged estimated changepoint $\overline{\tau}$ is quite similar in these models, with the earliest of 88.50 years of age and the latest 90.81 years. ApoE4 is associated with the mean of changepoint in all

the models with similar effect coefficient ($\beta_{\tau\_ApoE4} = 2.57 - 3.29$), suggesting a consistent correlation estimation. The model with polynomial regression function shows a tension zone as $\gamma = 4.0$ years, which matches the estimated width of the transition using bent-cable function. The slopes before and after the changepoint are pretty comparable in Broken-stick, bent-cable and polynomial model with that slope in phase 2 (after changepoint) is about 40-50 times faster than the slope in phase 1. The values of γ and β_{age} have no direct interpretation in Bacon-Watts and Griffith-Miller functions. The linking parameter ζ through changepoint is negatively associated with risk of AD and AD-free death in all models. Compared to the broken-stick model, the models with smooth transition function yield relatively smaller WAIC value. WAICs of Griffith-Miller and bent-cable models are the smallest among all suggesting a superior fit for the MAP data. From the simulation study, we know bent-cable has the best performance with smaller bias and better posterior interval coverage. Another reason that we prefer bent-cable function is its straightforward interpretation that Griffith-Miller does not provide.

Next we plot the estimated trajectories of cognitive decline according to different settings for the transition zone in Figure 13 for a selected MAP participant with AD diagnosed during the follow-up visit. All the models agree in terms of the time at changepoint, which is estimated between of an age of 93-94 years. Except for the broken-stick model which shows an abrupt change in direction, models with smooth transition function exhibit a gradual change over time. We notice the slope in the phase before changepoint is very flat in Bacon-Watts and is slightly slanting up in the Griffith-Miller model. The slight increase in cognition before dramatic decline takes over is not realistic in application of cognitive aging. The plots of bent-cable and polynomial are quite similar, and this agrees with the simulation analysis result from Chapter 4.

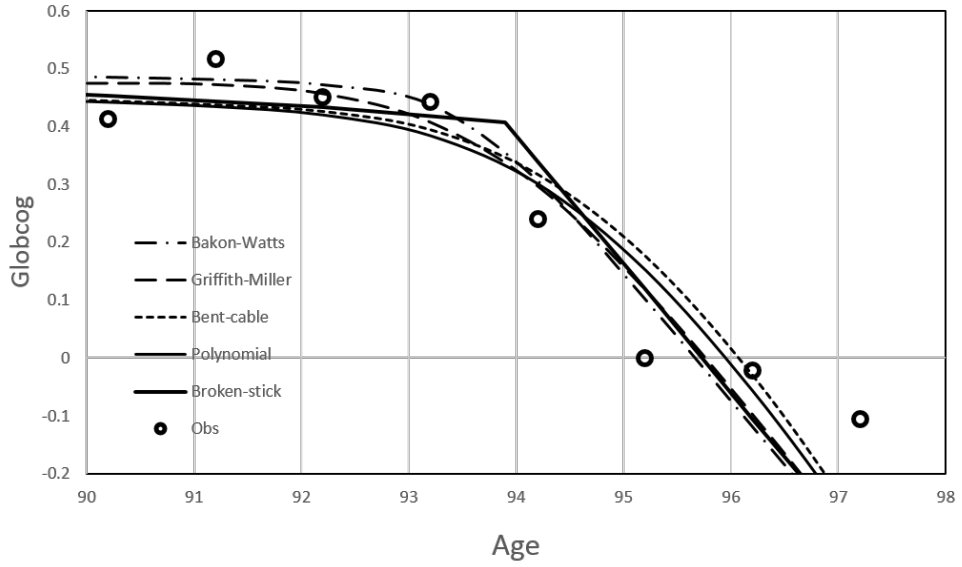

Figure 13: Estimated trajectory of cognitive decline according to different transition function for a MAP participant

5.2 Joint Model with Changepoint in Motor function Decline and Risk of AD in the Presence of Death

Now we would consider the non-cognitive function, such as motor, which may experience a change of decline with a break point during the progress to AD dementia. Motor function is a strong risk factor of AD and it was reported that change in motor performance such as gait speed may precede the onset of cognitive impairments and dementia (Buracchio et al., 2010). We next apply the joint model with quadratic bent-cable transition function for a changepoint on the motor function and this has never been investigated by this method before. We assume motor function exhibiting a similarly shaped trajectory as cognition, where it first decreases at a slow rate in a linear fashion, then goes through a curved transition phase, followed by a linear with an accelerated decreasing trend. For the baseline covariates in longitudinal submodel, we choose gender and body mass index (BMI) which are reported highly related with the motor skills in older adults (Xu et al., 2018). These two factors are also used in

predicting the mean of the age at changepoint that $\mu_{\tau_i} = \beta_{\tau_0} + \beta_{\tau_1} gender_i + \beta_{\tau_2} BMI\_bl_i$. The other settings of the model are the same as those in the model for cognitive function. Bayesian method is used for parameter estimation and we present results in Table XIX.

We find the acceleration in motor function decline appears at age of 81.6 (SD=4.5) years on average. BMI is negatively associated with the mean of the age at changepoint ($\beta_{\tau_2} = -2.34$), indicating motor decline is accelerated earlier for subjects with higher BMI value while gender has no effect. The difference of the slopes before and after transition is 0.020. The transition parameter γ is 1.7 years indicating a full length of about 3.5 year to finish the transition. The linking parameter ζs are both significant and negative in the survival function for AD and AD-free death. We also compare our proposed model fit (WAIC=-6585.5) with the model for motor function decline without considering a changepoint (WAIC=-5978.8) and the model using broken-stick function (WAIC=-6528.5). The WAIC-based criterion suggests that joint model with bent-cable function is supported by the data more than the other two in terms of fit. Our finding suggests the existence of a random changepoint in motor function decline before onset of AD or death and the change in decline rate is characterized as a gradual transition over time. We see for the AD cases (N=132) or AD-free death cases (N=200) in our analytical cohort, all these subjects underwent accelerated motor function decline, 9.2 (SD=2.3) years on average prior to AD diagnosis and 10.5 (SD=2.6) years before death. The result supports the hypothesis that motor impairment is an early marker for AD dementia. The acceleration in motor function decline can be used as useful information in the early detection of AD dementia.

To compare the changepoint in motor decline with that in cognitive decline, we take the difference of the two ages ($\tau_{cog_i} - \tau_{mot_i}$) within each subject and plot the histogram (Figure 14a). We find majority of the changepoints of motor are prior to the changepoint of cognition with mean difference of 7.4 (SD=3.8) years, and this is supported by previous research result that motor impairments precede cognitive impairment in predicting dementia

Table XIX: PARAMETER ESTIMATES OF JOINT ANALYSIS OF MOTOR FUNCTION DECLINE WITH BENT-CABLE TRANSITION FUNCTION FOR A CHANGEPOINT AND AD INCIDENCE WITH COMPETING RISK OF DEATH

Parameter	Post mean	SD	95%CI
Longitudinal			
Intercept (β_{00})	2.52	0.09	(2.33, 2.69)
Gender (β_{01})	0.08	0.02	(0.05, 0.12)
BMI (β_{02})	-0.11	0.01	(-0.13, -0.07)
Age (β_1)	-0.018	0.001	(-0.020, -0.015)
Age (β_2)	-0.020	0.02	(-0.023, -0.016)
Transition (γ)	1.71	1.29	(0.06, 4.59)
Ran_int (σ_0^2)	0.024	0.002	(0.020, 0.027)
Ran_cov (Cov_{01})	-0.002	0.000	(-0.002, -0.001)
Ran_age (σ_1^2)	0.000	0.000	(0.000, 0.000)
Error (σ_ϵ^2)	0.013	0.000	(0.012, 0.013)
Changepoint			
Intercept (β_{τ_0})	82.44	1.14	(80.18, 84.68)
Gender (β_{τ_1})	-1.28	1.13	(-3.52, 0.91)
BMI (β_{τ_2})	-2.34	0.96	(-4.25, -0.45)
Error (σ_τ^2)	46.06	6.79	(33.39, 59.76)
Survival			
Event=AD			
Edu (α_1)	0.04	0.04	(-0.04, 0.12)
ApoE4 (α_2)	0.52	0.28	(-0.03, 1.07)
changepoint (ζ)	-0.18	0.02	(-0.23, -0.14)
Ran_int (γ_1)	-2.69	1.30	(-5.20, -0.02)
Ran_age (γ_2)	-18.16	8.02	(-33.90, 2.82)
Event=death			
Gender (α_1)	0.14	0.25	(-0.38, 0.61)
Edu (α_2)	-0.03	0.03	(-0.08, 0.02)
changepoint (ζ)	-0.17	0.03	(-0.23, -0.11)
Ran_int (γ_1)	-2.42	0.97	(-4.14, -0.21)
Ran_age (γ_2)	-16.36	7.67	(-31.40, -1.46)

syndromes (Montero-Odasso et al., 2014. The scatter plot in Figure 14b shows a rough linear trend between the changepoints of the two functions. Their correlation coefficient is 0.66 indicating a strong positive association.

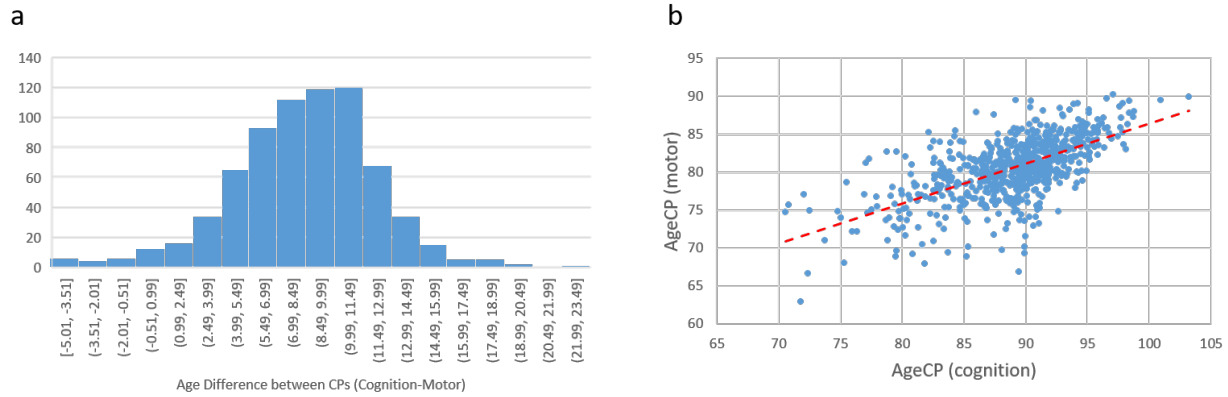

Figure 14: Changepoint comparison between cognition and motor function
a. Histogram plot for the age difference between changepoints ($\tau_{cog_i} - \tau_{mot_i}$)
b. Scatter plot of the age at changepoint for cognition and motor function

5.3 Joint Model for Cognition and Motor Function Decline with Random Changepoint and Risk of AD in the Presence of Death

Motivated by the results from the separate analysis of the changepoint in cognition and motor function degradation, we next use the joint model with bivariate changepoints. The purpose of the bivariate changepoint model is to provide a useful framework to assess the relationship between the two changepoints by taking their correlation structure into account. It is interesting to investigate the longitudinal trajectories of cognition and motor function simultaneously and determine whether the change in motor decline proceeds cognitive impairment or vice versa.

5.3.1 Model Specification

Based on the simulation study in Chapter 4, we see the good performance of the joint model for bivariate longitudinal outcomes with random changepoint under bent-cable transition structure. We fit the proposed joint model of bivariate changepoint to MAP cohort using the previously described Bayesian framework. Specifically, we use age as the timescale. For each longitudinal outcome, we keep the same longitudinal modeling structure as in their separate model except that we only keep ApoE4 and BMI as the covariates in the regression function for the mean age at changepoint of cognitive decline and motor decline, respectively, because they show significant association with the changepoint in previous analysis.

$$\begin{cases} Globcog_{ij} = (\beta_{100} + \beta_{101} gender_i + \beta_{102} edu_i + b_{10i}) + \beta_{11} age_{ij} + \\ \qquad\qquad (\beta_{12} + b_{11i}) q_{bent\_cable}(age_{ij}; \tau_{1i}, \gamma_1) + \epsilon_{1ij}, \\ Globmot_{ij} = (\beta_{200} + \beta_{201} gender_i + \beta_{202} BMI_i + b_{20i}) + \beta_{21} age_{ij} + \\ \qquad\qquad (\beta_{22} + b_{21i}) q_{bent\_cable}(age_{ij}; \tau_{2i}, \gamma_2) + \epsilon_{2ij}, \end{cases} \tag{5.3}$$

The two longitudinal outcomes are then correlated through the random effects

$$\begin{pmatrix} b_{cogi} \\ b_{moti} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, D = \begin{pmatrix} \Sigma_{bcog} & \Sigma_{bcog\_mot} \\ \Sigma_{bcog\_mot}^{\top} & \Sigma_{bmot} \end{pmatrix} \right). \tag{5.4}$$

and their changespoints

$$\begin{pmatrix} \tau_{cogi} \\ \tau_{moti} \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_{\tau_{cog0}} + \beta_{\tau_{cog1}} ApoE4_i \\ \beta_{\tau_{mot0}} + \beta_{\tau_{mot1}} BMI_i \end{pmatrix}, D = \begin{pmatrix} \sigma^2_{\tau_{cog}} & \rho\sigma_{\tau_{cog}}\sigma_{\tau_{mot}} \\ \rho\sigma_{\tau_{cog}}\sigma_{\tau_{mot}} & \sigma^2_{\tau_{mot}} \end{pmatrix} \right). \tag{5.5}$$

Our particular interest is the relationship between the changepoints of cognition and motor function measures. For the linking structure between longitudinal and survival processes, only the parameters (τ_{1i}, b_{10i} and b_{11i}) based on longitudinal submodel of cognition are considered in AD and death hazard functions. We have several reasons to choose this setting. First, based on model selection result in Chapter 3 (Table VI) for the shared random effects joint model of cognition and motor function, we see the random effects from cognition longitudinal submodel are sufficient to describe the dependence between the longitudinal trajectories and rate of AD incidence. Second, the changepoint of motor function appears to be highly correlated with that in cognitive decline and keeping one changepoint (cognition) will avoid the multicollinearity problem. The strong association between the time point of the acceleration in cognitive decline and risk of AD or death has been proved by previous studies. Third, when we compare the model with different linking structure (Table XX), sharing parameters from longitudinal model of cognition in both survival functions yields the smallest WAIC (-10305.6) and thus is preferred. Last, as the model structure is already complex, we try to avoid using too many parameters in the model for a good estimation efficiency.

Table XX: MODEL SELECTION FOR JOINT MODELS WITH DIFFERENT SHARED PARAMETERS

Model	AD survival function	Death survival function	WAIC
I	$\zeta_{AD}\tau_{cog} + r_{AD1}b_{0cog} + r_{AD2}b_{1cog}$	$\zeta_{D}\tau_{cog} + r_{D1}b_{0cog} + r_{D2}b_{1cog}$	-10305.6
II	$\zeta_{AD}\tau_{mot} + r_{AD1}b_{0mot} + r_{AD2}b_{1mot}$	$\zeta_{D}\tau_{mot} + r_{D1}b_{0mot} + r_{D2}b_{1mot}$	-10171.1
III	$\zeta_{AD}\tau_{cog} + r_{AD1}b_{0cog} + r_{AD2}b_{1cog}$	$\zeta_{D}\tau_{mot} + r_{D1}b_{0mot} + r_{D2}b_{1mot}$	-10078.6
IV	$\zeta_{AD}\tau_{mot} + r_{AD1}b_{0mot} + r_{AD2}b_{1mot}$	$\zeta_{D}\tau_{cog} + r_{D1}b_{0cog} + r_{D2}b_{1cog}$	-10280.7

We use Bayesian MCMC approach to obtain the estimators for the proposed joint models in WinBUGS. The prior distributions are similar to those in the simulation study. Since the model is complex and contains more covariates, we consider longer burn-in time and monitor the trace and density plots of MCMC interactions to check the convergence behaviors.

5.3.2 Result

Table XXI shows the summary statistics of the posterior distributions of parameters in the model. The mean slope of cognitive decline is -0.005 points per year before changepoint, -0.294 after changepoint; the mean slope before the changepoint is -0.018 and is -0.043 after the changepoint for motor function decline. These parameter estimates are pretty similar to those obtained from separate models and the slope change is more dramatic in cognitive decline compared to that in motor function. The time of the acceleration in cognitive decline comes early for ApoE4 carriers ($\beta = -2.39$) while lower BMI delays the age of changepoint in motor decline ($\beta = -0.69$), but its effect is only marginally significant. The transition zone for cognition and motor function decline is 4.4 years and 5.1 years, respectively. The posterior mean of the correlation coefficient between the two changepoints is 0.89 (0.82, 0.92), suggesting that the change in cognitive decline is positively and highly correlated with the change in motor function decline. The mean age of changepoint is 88.7 (SD=5.1) years for cognitive function and 83.2 (SD=6.2) years for motor function. Furthermore, we find the acceleration in motor function decline occurs in years before the acceleration of cognitive decline for 99.4% (N=713) of the sample and the averaged within-subject difference ($\tau_{cogi} - \tau_{moti}$) is 5.6 (SD=2.3) years.

We also run the model in which no correlation between the two changepoints is assumed ($\rho = 0$). Though the estimates for the interested parameters are similar, the reduced model yields a much bigger WAIC (-8612.5) compared to the model considering changepoint correlation (WAIC=-10305.6).

The linking parameters for the changepoint ζ between the longitudinal and survival processes is negative ($\zeta_{AD} = -0.74$ and $\zeta_{Death} = -0.42$) and significant in both AD and Death survival functions. All the shared random effects terms are significant ($r_{1\_AD} = -1.92$, $r_{2\_AD} = -5.62$ and $r_{2\_Death} = -9.47$) except for the random intercept for death risk. These

Table XXI: PARAMETER ESTIMATES OF JOINT ANALYSIS OF COGNITION AND MOTOR FUNCTION DECLINE WITH BENT-CABLE TRANSITION FUNCTION AND AD INCIDENCE WITH COMPETING RISK OF DEATH

Parameter	Post mean	SD	95%CI
Longitudinal			
Outcome=Globcog			
Intercept (β_{00})	0.07	0.10	(-0.27, 0.12)
Gender (β_{01})	-0.10	0.05	(-0.18, -0.03)
Educ (β_{02})	0.06	0.00	(0.05, 0.07)
Age (β_1)	-0.005	0.001	(-0.007, -0.003)
Age (β_2)	-0.29	0.01	(-0.31, -0.27)
Transition (γ)	2.22	0.18	(1.87, 2.56)
Ran_int (σ_0^2)	0.125	0.007	(0.112, 0.139)
Ran_cov (Cov_{01})	0.005	0.004	(-0.003, 0.013)
Ran_age (σ_1^2)	0.037	0.004	(0.030, 0.045)
Error (σ_ϵ^2)	0.03	0.00	(0.03, 0.03)
Outcome=Globcmot			
Intercept (β_{00})	2.56	0.06	(2.44, 2.68)
Gender (β_{01})	0.07	0.01	(0.04, 0.09)
BMI (β_{02})	-0.10	0.01	(-0.12, -0.07)
Age (β_1)	-0.018	0.001	(-0.020, -0.017)
Age (β_2)	-0.025	0.002	(-0.028, -0.022)
Transition (γ)	2.54	1.26	(0.21, 4.84)
Ran_int (σ_0^2)	0.023	0.002	(0.020, 0.026)
Ran_cov (Cov_{01})	-0.002	0.000	(-0.002, -0.001)
Ran_age (σ_1^2)	0.000	0.000	(0.000, 0.000)
Error (σ_ϵ^2)	0.01	0.00	(0.01, 0.01)
Cov_{cog0_mot0}	0.010	0.002	(0,005, 0.015)
Cov_{cog0_mot1}	-0.002	0.000	(-0.003, -0.001)
Cov_{cog1_mot0}	0.008	0.002	(0.003, 0.012)
Cov_{cog1_mot1}	0.001	0.000	(0.000, 0.001)
Changepoint (cognition)			
Intercept (β_{0_τ})	89.09	0.30	(88.54, 89.66)
ApoE4 (β_{1_τ})	-2.39	0.42	(-3.31, -1.66)
Error (σ_τ^2)	40.55	2.04	(36.25, 44.41)
Changepoint (motor)			

TABLE XXI: PARAMETER ESTIMATES OF JOINT ANALYSIS OF COGNITION AND MOTOR FUNCTION DECLINE WITH BENT-CABLE TRANSITION FUNCTION AND AD INCIDENCE WITH COMPETING RISK OF DEATH (CONTINUED)

Parameter	Post mean	SD	95%CI
Intercept (β_{0_τ})	84.29	0.42	(83.38, 85.04)
BMI (β_{1_τ})	-0.69	0.43	(-1.45, 0.02)
Error (σ^2_τ)	60.73	3.33	(55.34, 68.31)
correlation (ρ)	0.89	0.03	(0.87, 0.92)
Survival(Event=AD)			
ApoE4 (α)	0.16	0.24	(-0.34, 0.62)
Changepoint (ζ)	-0.74	0.03	(-0.79, -0.69)
Ran_int (r_1)	-1.92	0.40	(-2.66, -1.07)
Ran_age (r_2)	-5.62	0.62	(-6.75, -4.32)
Survival(Event=Death)			
Gender (α)	0.42	0.22	(0.02, 0.89)
Changepoint (ζ)	-0.42	0.02	(-0.47, -0.37)
Ran_int (r_1)	0.28	0.33	(-0.35, 0.92)
Ran_age (r_2)	-9.47	0.95	(-11.34, -7.51)

Table XXII: MODEL COMPARISON FOR JOINT MODELS WITH OR WITHOUT CHANGE-POINT CORRELATION

		Cognition				Motor			
Model	WAIC	$\tilde{\tau}$	γ_{CP}	β_1	β_2	$\tilde{\tau}$	γ_{CP}	β_1	β_2
ρ=0	-8612.5	88.6	2.0	-0.006	-0.289	80.7	1.5	-0.017	-0.035
$\rho \neq 0$	-10305.6	88.7	2.2	-0.005	-0.298	83.2	2.5	-0.018	-0.043

results suggest the shape of the cognition trajectory across all time is strongly linked to the risk of AD and death risk is more related to the time when the acceleration in decline starts and the slope after. Of other covariates we put in the proportional hazard function, ApoE4 increases the risk of AD, but its effect is not significant; Men have significantly higher risk of death ($\alpha = 0.42$) compared to women.

Figure 15 shows the smoothed 3-D distribution of the random changepoints of cognition and motor function estimated (a) from two separate joint model with AD and death and (b) simultaneously from one joint model with AD and death. Both plots show a strong correlation between the two changepoints. The surface is much smoother, and the peak is very narrow and steep for the joint model with bivariate longitudinal outcomes.

5.4 Discussion

In the last chapter, the proposed methodology is demonstrated with application to the MAP cohort. We jointly described the evolution of cognition and risk of AD dementia taking into account a change in the linear trend in the longitudinal marker trajectory. The choice of a smooth curve for the transition depends on the nature of the data. We adopt the bent-cable regression method, which models the transition as a quadratic phase with unknown width. It is desirable to formulate a model that can describe the transition zone accurately and also provides meaningful and interpretable estimates.

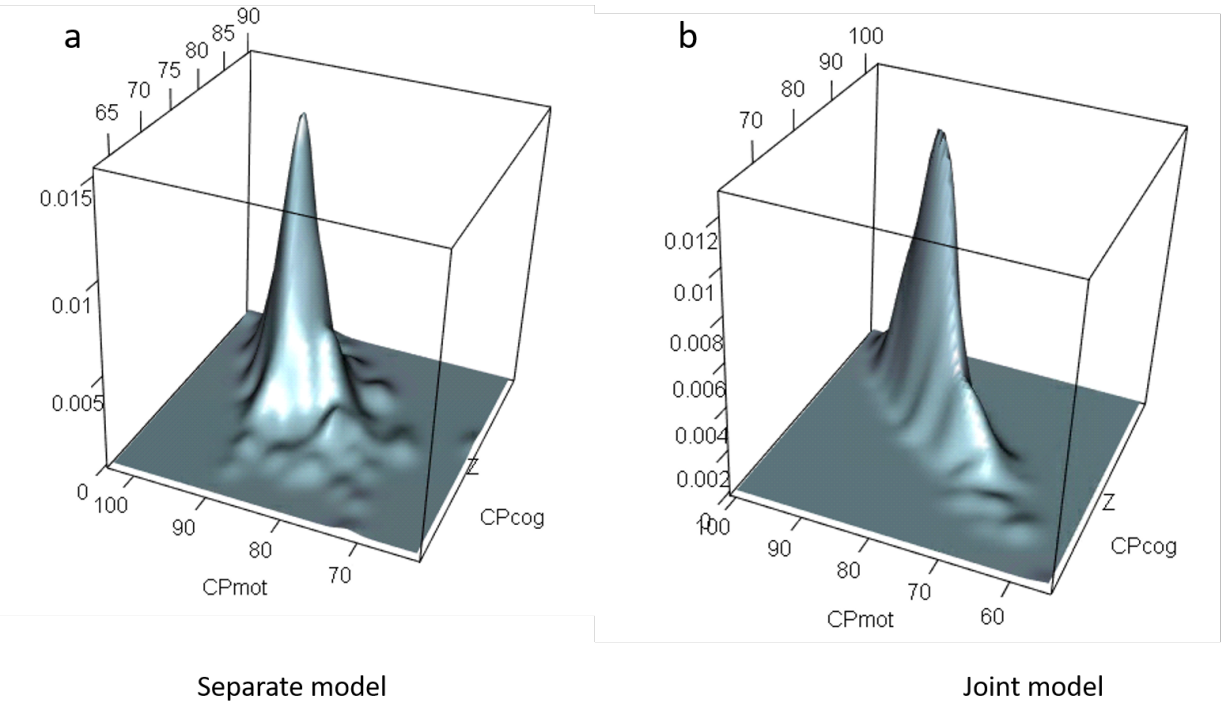

Figure 15: 3D distribution of changepoints in cognition and motor function decline

5.4.1 Implications of Results

The estimated age at changepoint in cognitive decline starts as early as 71 years old and mean is 89. The average length before the diagnosis of AD for the changepoint is 4.0 years and is 2.1 years before death without AD diagnosed. If we also account for the death with AD diagnosed, the average length is 4.5 year before death for the changepoint. These results are close to what was reported previous research that Hall et al., 2000, found the changepoint happened on average 5.1 years before diagnosis of dementia and Amieva et al., 2005, showed a dramatic increase in the rate of decline about 3 years before the diagnosis. Moreover, Wilson et al., 2003, found abrupt change in the rate of cognitive decline 3 to 6 years before death. Later, Hout et al., 2010, estimated the changepoint around 6 years before death. Most recent review article based on systematic search of 35 studies for cognitive performance and neurology outcome measures preceding dementia or death summarized that

change for cognitive function ranged from 1 to 11 years before dementia onset time, and 3 to 15 years prior to death (Karr et al., 2018).

We reported the changepoint in cognitive decline is related to ApoE4 ($\beta = -3.06$) but not education level. Previous studies found changepoint was delayed for participants with more education (Yu and Ghosh, 2010; Jacqmin-Gadda et al., 2006), while other research reported there was no education-related difference for the changepoint (Li et al., 2015). MAP cohort has relatively higher education level compared to other cohorts and only a small portion of the sample have less than 9 years of education. This could be a reason for not detecting a difference in changepoint time by education. There was only one study that examined the effect of genetic risk of ApoE4 in the changepoint model among dementia cases (Li et al., 2015), however, they found the ApoE4 allele did not affect the location of the changepoint before AD. Another study identified an earlier changepoint for ApoE4 carriers by roughly 9 month in cognitive functioning prior to death (Yu et al., 2013). Inconsistent findings across studies in terms of changepoint component estimation could be due to discrepant research designs, missing data and different analysis methods. Most previous analysis on changepoint in cognitive decline only focused on case cohort and formulation of transition for an abrupt change was commonly used. Our model overcomes these drawbacks by taking into account the health status as AD dementia and death to avoid selection bias. In addition, the model with smooth transition possesses great flexibility about the shapes of changepoint trajectories and improves the model fitting to the data.

This is the first time that motor function trajectory is modeled for a changepoint. The data support the hypothesis that persons who will develop AD, begin to experience accelerated motor function decline many years before their diagnosis. Motor function may serve as an early predictor of a future AD diagnosis. However, we notice the increase in the rate of decline (2-3 times) in motor is not as dramatic as in cognitive decline (about 30 fold), and the difference may not be easily detected in clinical practice.

In the final application, we modeled the trajectories of cognition and motor function for both their changepoints taking account of their correlation, and introducing the longitudinal trajectories to the AD/death competing risks model. The analysis of bivariate outcomes having different changepoints has been carried out previously (Hall et al., 2001; Yang and Gao, 2013; Segalas et al., 2020). These analysis only focused on the AD dementia cases. Moreover, the correlation structures of the two longitudinal markers are too simple and may not capture the whole association. Some work did not use smooth transition to model the changepoint or did not have obvious interpretation for the changes. Here, we provide a unified framework that accounts for all the issues together, in the same time, performs the comparison of the changepoints between the longitudinal markers.

5.4.2 Limitation

Joint modeling approach is the appropriate statistical tool for assessing the progression of longitudinal markers accounting for the health statuses that are related with the endpoint or vice versa. Identification of a changepoint would provide evidence for an acceleration in trend and its location would be helpful to determine the timecourse for early detection and treatment of the disease. Since it is a new and fast developing field in biostatistics, there are many limitations in this thesis work that we like to mention.

First, changepoint studies may have different settings leading to different results with respect to study design, statistical model, inclusion of covariates. As noticed, not many covariates are considered in our model especially the interactions of time with the covariates due to the joint model complexity and computation burden. Ignoring or under-use of potential variable information may bias the results and conclusions.

Second, for modeling purposes, we only include participants with at least 5 measurements of cognition and motor function for our analysis. This setting makes the analysis inevitably

conditional on the subjects who are free of events for a relatively long period of time during follow-up. This selection could lead to biased results as it limits the investigation on a subset of population healthier than the rest of the cohort.

Third, our joint model assumes that every subject develops AD dementia eventually and the trajectories of aging function markers consist of a changepoint. Although there is a rapid growth in the case number each year, the prevalence of dementia is 5-8% in people aged 60+ years. In reality, there are evidence that some people are immune to AD and it was reported that individuals exist whose brains are devoid of disease even at the age of 90 years and beyond (Cooner et al., 2018). The standard survival model is assuming that all the subjects will experience the event eventually when the follow-up is as long as needed. But in the presence of a dementia free subgroup, standard survival model may not be appropriate. It could be expected that a fraction of people, especially among non-AD cases, would present a linear trend of cognition only. Thus, a joint model, where a cure rate is taken to separate the subjects into one group having risk of AD with accelerated rate of cognitive decline and another group of null risk with linear cognitive trend, should be more realistic. That is the underlying model assumes a mixture distribution for the changepoint. If this null risk portion exists, Kaplan-Meier survival curve of AD should exhibit a plateau. However, in MAP data such plateau is not observed, and the null risk portion is not applicable and a cure model does not work for our study (see Appendix A for Joint model with cure fraction).

5.4.3 Future Work

It is possible that there exists a second changepoint for another deterioration in cognition just round the time of AD diagnosis or death. Estimation of mixed models with multiple random changepoints could offer more realistic descriptions of late-life cognitive change but would be computationally challenging.

In the late-life period, the trajectories of cognition are heterogeneous as described in Chapter 3. This is the reason for that random (subject-specific) changepoint fits better than a fixed one. A single set of random variables (with Gaussian distribution) modeling the correlation between the repeated measures of longitudinal markers and their relationship with the event risk may not be sufficient. As the pattern of age-health trajectories is highly variable and several longitudinal studies of AD dementia report the heterogeneity in the long-term progression, characterized by the patients' complex cognitive evolution (Goyal et al., 2018), the link between the longitudinal outcomes and events needs to be more precisely defined. A useful alternative would be to jointly model the cognitive decline and risk of AD using a latent stochastic process but introducing the changepoint within each latent class. This might lead to a heavy computation and possibility of unidentifiable latent classes, especially when we consider a smooth transition to model the changepoint. One option is that we can artificially designate fixed numbers of classes, for example, two latent classes with one having an accelerated rate of cognitive decline and another with linear cognitive trend. This approach will release the assumption that all subjects have a changepoint. The framework should be carefully designed when multiple outcomes are introduced in the joint model.

In this study we only focus on cognitive and motor change in the pre-AD dementia phase, and the measures after the diagnosis are not used. It is also interesting to learn the cognition and motor trends after dementia diagnosis and this will help to understand the whole picture of the disease development.

CITED LITERATURE

Abramowitz, M. and Stegun, I. (1972). "Handbook of mathematical functions with formulas, graphs, and mathematical tables". 10th printing. New York: Dover, pp. 12–35.

Alhurani, RE., Vassilaki, M., Aakre, J., Mielke, MM., Kremers, WK., Machulda, MM., Geda, YE., Knopman, DS., Peterson, RC., and Roberts, RO. (2016). "Decline in weight and incident mild cognitive impairment: Mayo clinic study of aging". *JAMA neurology* 7.21, pp. 439 –446.

Alzheimer's_Association (2019). "2019 Alzheimer's Disease Facts and Figures". *Alzheimers Dement* 15.3. `https://www.alz.org/media/documents/alzheimers-facts-and-figures-2019-r.pdf`, 321–387.

Amieva, H., Goff, M., Millet, X., Orgogozo, JM., Peres, K., Barberger-Gateau, P., Jacqmin-Gadda, H., and Dartigues, JF. (2008). "Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms". *Annals of neurology* 64.5, pp. 492–498.

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J., Carret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., and Dartigues, J. (2005). "The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study". *Brain: a journal of neurology* 128.5, pp. 1093–1101.

Amieva, H., Mokri, H., Goff, ML., Meillon, C., Jacqmin-Gadda, H., Foubert-Samier, A., Orgogozo, J., Stern, Y., and Dartigues, J. (2014). "Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: a study of 20 years of cognitive decline". *Brain: A journal of neurology* 137.4, pp. 1167–1175.

Andrinopoulo, ER., Nasserinejad, K., Szczesniak, R., and Rizopoulos, D. (2020). "Integrating latent classes in the Bayesian shared parameter joint model of longitudinal and survival outcomes". *Statistical methods in medical research* 29.11, pp. 3294–3307.

Asparouhov, T. and Muthen, B. (2011). "Using Bayesian priors for more flexible latent class analysis". Section on government statistics, the Joint Statistical Meetings.

Asparouhov, T. and Muthén, B. (2014). "Auxiliary variables in mixture modeling: three-step approaches using Mplus". *Structural equation modeling A multidisciplinary journal* 21.3, pp. 329–341.

Bacon, DW. and Watts, DG. (1971). "Estimating the transition between Two intersecting straight lines". *Biometrika* 58.3, pp. 525–534.

Bennett, DA., Buchman, AS., Boyle, PA., Barnes, LL., Wilson, RS., and Schneider, JA. (2018). "Religious Orders Study and Rush Memory and Aging Project". *Journal of Alzheimer's disease* 64.s1, S161–S189.

Bennett, DA., Schneider, JA., AS., AS. Buchman, Barnes, LL., Boyle, PA., and Wilson, RS. (2012). "Overview and findings from the rush memory and aging project". *Current Alzheimer research* 9.6, pp. 646–663.

Berkson, J. and Gage, RP. (1952). "Survival curve for cancer patients following treatment". *Journal of the American statistical Association* 47.259, pp. 501–515.

Berlin, KS., William, NA., and Parra, GR. (2014). "An introduction to latent variable mixture modeling: Overview and cross-sectional latent class and latent profile analyses". *Journal of pediatric psychology* 39.9, pp. 174–187.

Buchman, A. and Bennett, DA. (2011). "Loss of motor function in preclinical Alzheimer's disease". *Expert review of neurotherapeutics* 11.5, pp. 665–676.

Buracchio, T., Dodge, H., and Howieson, D. (2010). "The Trajectory of gait speed preceding mild cognitive impairment". *JAMA neurology* 67.8, pp. 980–986.

Carlin, B., Gelfand, A., and Smith, A. (1992). "Hierarchical Bayesian analysis of changepoint problems". *Journal of the royal statistical society* 41.2, pp. 389–405.

Celeux, G., Hurn, M., and Robert, PC. (2000). "Computational and inferential difficulties with mixture posterior distributions". *Journal of the American statistical Association* 95.45, pp. 957–970.

Chang, CC., Zhao, Y., Lee, CW., and Ganguli, M. (2012). "Smoking, death, and Alzheimer's disease: A case of competing risks". *Alzheimer disease and associated disorders* 26.4, pp. 300–306.

Chibnik, LB., Wolters, FJ., Backamn, K., Beiser, A., and al., et (2017). "Trends in the incidence of dementia: design and methods in the Alzheimer Cohorts Consortium". *European journal of epidemiology* 32.10, pp. 931–938.

Chiu, G., Lockhart, R., and Routledge, R. (2006). "Bent-cable regression theory and applications". *Journal of the American statistical association* 101, pp. 542–553.

Cho, SJ., Cohen, AS., and Kim, SH. (2013). "Markov chain Monte Carlo estimation of a mixture item response theory model." *Journal of statistical computation and simulation* 38.2, pp. 278–306.

Cooner, F., Banerjee, S., Carlin, BP., and Sinha, D. (2018). "Flexible cure rate modeling under latent activation schemes". *Journal of the American statistical Association* 102.478, pp. 560–572.

Cox, DR. (1972). "Regression models and life-tables". *Journal of the royal statistical society* 34.2, pp. 187–220.

Cullen, B., O'Neill, B., Evans, JJ., Coen, RF., and Lawlor, BA. (2007). "A review of screening tests for cognitive impairment". *Journal of neurology, neurosurgery and psychiatry* 78.8, pp. 790–799.

Dantan, E., Joly, P., Dartigues, JF., and Jacqmin-Gadda, H. (2011). "Joint model with latent state for longitudinal and multistate data". *Biostatistics* 12.4, pp. 723–736.

Dartigues, JF., Gagnon, M., Michel, P., Letenneur, L., Commenges, D., Barberger-Gateau, P., and al, et (1991). "The Paquid research program on the epidemiology of dementia. Methods and initial results". *Revue neurologique* 147.3, pp. 225–230.

Dessiso, AH. and Goshu, AT. (2017). "Bayesian joint modelling of longitudinal and survival data of HIV/AIDS Patients: A case study at Bale Robe General Hospital, Ethiopia". *American journal of theoretical and applied statistics* 6.64, pp. 182–190.

Faucett, CL. and Thomas, DC. (1996). "Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach". *Statistics in medicine* 15.15, pp. 1663–1685.

Fine, JP. and Gray, RJ. (1999). "A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association". *Journal of the American statistical Association* 94, pp. 496–509.

Formann, AK. (2011). "Constrained latent class models: Theory and applications". *British journal of mathematical and statistical psychology* 38.1, pp. 87–111.

Ganjalia, M. and Baghfalaki, T. (2015). "A copula approach to joint modeling of longitudinal measurements and survival times using MCEM with application to AIDS studies". *Journal of biopharmaceutical statistics* 25.5, pp. 1077–1099.

Gao, S. (2004). "A shared random effect parameter approach for longitudinal dementia data with non-ignorable missing data". *Statistics in medicine* 23.2, pp. 211–219.

Gelman, A. and Rubin, DB. (1992). "Inference from iterative simulation using multiple sequences". *Statistical science* 7, 457–511.

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to calculating posterior moments". *Bayesian statistics* 4. Clarendon Press, Oxford, UK, pp. 169–193.

Goyal, D., Tjandra, D., Migrino, RQ., Giordani, B., Syed, Z., and Wiens, J. (2018). "Characterizing heterogeneity in the progression of Alzheimer's disease using longitudinal clinical and neuroimaging biomarkers". *Diagnostic assessment & Prognosis* 10, pp. 629–637.

Grambsch, PM. and Therneau, TM. (1994). "Proportional hazards tests and diagnostics based on weighted residuals". *Biometrika* 81.3, pp. 515–526.

Griffiths, D. and Miller, A. (1973). "Hyperbolic regression - a model based on two-phase piecewise linear regression with a smooth transition between regimes". *Communications in statistics* 2, pp. 561–569.

Hall, C., Ying, J., Kuo, L., and Lipton, R. (2003). "Bayesian and profile likelihood change point methods for modeling cognitive function over time". *Computational statistics Data analysis* 43.1, pp. 91–109.

Hall, CB., Lipton, RB., Sliwinski, M., and Stewar, WF. (2000). "A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease". *Statistics in medicine* 19.11-12, pp. 1555–1566.

Hall, CB., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M., and Lipton, RB. (2001). "Estimation of bivariate measurements having dierent change points, with application to cognitive ageing". *Statistics in medicine* 20.24, pp. 3695–3714.

Harrison, X., Donaldson, L., Correa-Cano, M., Evans, J., Fisher, D., Goodwin, C., Robinson, B., Hodgson, D., and Inger, R. (2018). "A brief introduction to mixed effects modelling and multi-model inference in ecology". *PeerJ* 6, e4794.

Hashemi, R. and Jacqmin-Gadda, H. (2003). "A latent process model for joint modeling of events and marker". *Lifetime data analysis* 9.4, pp. 331–343.

Hawkins, D., Allen, D., and Stromberg, A. (2001). "Determining the number of components in mixtures of linear models". *Computational Statistics & Data Analysis* 38.1, pp. 15–48.

Helzner, EP., Scarmeas, N., Cosentino, S., Tang, MX., Schupf, N., and Stern, Y. (2008). "Survival in Alzheimer disease". *Neurology* 71.19, pp. 1489–1495.

Henderson, R., Diggle, P., and Dobson, A. (2000). "Joint modelling of longitudinal measurements and event time data". *Biostatistics* 1.4, pp. 465–480.

Hernan, MA., Alonsoa, A., and Logroscino, G. (2008). "Cigarrete smoking and dementia: potential selection bias in the elderly". *Epidemiology* 19.3, pp. 448–450.

Hester, RL., Kinsella, GJ., Ong, B., and McGregor, J. (2005). "Demographic influences on baseline and derived scores from the trail making test in healthy older Australian adults". *The Clinical neuropsychologist* 19.1, pp. 45–54.

Hinkley, D. (1970). "Inferences about the change-point in a sequence of random variables". *Biometrika* 57.1, pp. 1–17.

Horrocks, J. and Heuvel, M. (2009). "Prediction of pregnancy: A joint model for longitudinal and binary data". *Bayesian analysis* 4, pp. 523–538.

Hout, A. van den, Muniz-Terrera, G., and Matthews, FE. (2010). "Smooth random change point models". *Statistics in medicine* 30.6, pp. 599–610.

Ibrahim, JG., Chu, H., and Chen, LM. (2010). "Basic concepts and methods for joint models of longitudinal and survival data". *Journal of clinical oncology* 28.16, pp. 2796–2801.

Jacqmin-Gadda, H., Commenges, D., and Dartigues, J. (2006). "Random changepoint model for joint modeling of cognitive decline and dementia". *Biometrics* 62.1, pp. 254–260.

Jacqmin-Gadda, H., Proust-Lima, C., Taylor, JMG., and Commenges, D. (2010). "Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model". *Biometrics* 66.1, pp. 11–19.

Jansen, WJ., Wilson, RS., Visser, PJ., Nag, S., Schneider, JA., James, BD., Leurgans, SE., Capuano, AW., Bennett, DA., and Boyle, PA. (2018). "Age and the association of dementia-related pathology with trajectories of cognitive decline". *Neurobiology of aging* 61, pp. 138–145.

Karr, JE., Graham, RB., Hofer, SM., and Muniz-Terrera, G. (2018). "When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death". *Psychology and aging* 33.2, pp. 195–218.

Khan, SA. and Kar, SC. (2017). "Generalized bent-cable methodology for changepoint data: a Bayesian approach". *Journal of applied statistics* 45, pp. 1799–1812.

Kleinbaum, DG. and Klein, M. (2012). "Competing risks survival analysis". *Survival analysis: A self-learning text*. Book, pp. 425–495.

Kueper, J., Speechley, M., Lingum, N., and Montero-Odasso, M. (2017). "Motor function and incident dementia: a systematic review and meta-analysis". *Age and ageing* 46.5, pp. 729–738.

Kukull, WA., Higdon, R., Bowen, JD., McCormick, WC., Teri, L., Schellenberg, GD., Belle, G. van, Jolley, L., and Larson, EB. (2002). "Dementia and Alzheimer disease incidence: A prospective cohort study". *Archives of neurology* 59.11, pp. 1737–1746.

Kumar, A., Sidhu, J., Goyal, A., and Tsao, JW. (2020). "Alzheimer disease". *StatPearls [Internet].Treasure Island (FL): StatPearls Publishing*.

Kuo, Y., Li, C., Sung, J., Chang, C., Wang, J., Sun, C., Wu, J., and Chang, Y. (2019). "Risk of dementia in patients with end-stage renal disease under maintenance dialysis - a nationwide population-based study with consideration of competing risk of mortality". *Alzheimer's research & Therapy* 11.1, pp. 1–12.

Lawrence, A., Boye, M., Crowther, M., Ibrahim, JG., Quartey, G., Micallef, S., and Bois, F. (2015). "Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group". *Statistics in medicine* 34.14, pp. 2181–2195.

Lazarsfeld, PF. (1950). "The logical and mathematical foundation of latent structure analysis. In S. Stouffer (Ed.)," *The American soldier* 4, pp. 413–472.

Li, C. (2016). "The Fine-Gray model under interval censored competing risks data". *Journal of multivariate analysis* 143, pp. 327–344.

Li, C., Dowling, NM., and Chappell, R. (2015). "Quantile regression with a change-point model for longitudinal data: An application to the study of cognitive changes in preclinical alzheimer's disease". *Biometrics* 71.3, pp. 625–35.

Li, K. and Luo, S. (2019). "Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease". *Statistical methods in medical research* 28.2, pp. 327–342.

Li, S., Zheng, M., and Gao, S. (2017). "Joint modeling of longitudinal cholesterol measurements and time to onset of dementia in an elderly African American Cohort". *Biostatistics & Epidemiology* 1, pp. 148–160.

Lin, H., Turnbull, BW., McCulloch, C., and Slate, E. (2002). "Latent class models for joint analysis of longitudinal biomarker and event process data: Application to longitudinal prostate-specific antigen readings and prostate cancer". *Journal of the American statistical Association* 97, pp. 53–65.

Lundervold, A., Wollschläger, D., and Wehling, E. (2014). "Age and sex related changes in episodic memory function in middle aged and older adults". *Scandinavian journal of psychology* 55.3, pp. 225–232.

Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). "The BUGS project: Evolution, critique and future directions". *Statistics in medicine* 28.25, 3049—3067.

Lunn, DJ., Thomas, A., Best, N., and Spiegelhalter, D. (2000). "WinBUGS – a Bayesian modelling framework: Concepts, structure, and extensibility". *Statistics and computing* 10, pp. 325–337.

Lv, X., Li, W., Ma, Y., Chen, H., Zeng, Y., Yu, X., Hofman, A., and Wang, H. (2019). "Cognitive decline and mortality among community-dwelling Chinese older people". *BMC Medicine* 17.63, pp. 507–514.

Marquardt, D. (1963). "An algorithm for least squares estimation of nonlinear parameters". *Journal of the society for industrial and applied mathematics* 11.2, pp. 431–441.

McCrink, L., Marshall, A., and Cairns, K. (2003). "Advances in joint modelling: A review of recent developments with application to the survival of end stage renal disease patients". *International statistical review* 81.2, pp. 249–269.

Mokhles, MM., Rizopoulos, D., Andrinopoulou, ER., Bekkers, JA., Roos-Hesselink, JW., Lesaffre, E., Bogers, AJ., and Takkenberg, JJ. (2012). "Autograft and pulmonary allograft performance in the second post-operative decade after the Ross procedure: insights from the Rotterdam prospective cohort study". *European heart journal* 33, pp. 2213–2224.

Montero-Odasso, M., Oteng-Amoako, A. fua, Speechley, M., Gopaul, K., Beauchet, O., Annweiler, C., and Muir-Hunter, SW. (2014). "The motor signature of mild cognitive impairment: Results from the gait and brain study". *The journals of gerontology. Series A, Biological sciences and medical sciences* 69.11, pp. 1415–1421.

Morris, MC., Tangney, CC., Wang, Y., Sacks, FM., Bennett, DA., and Aggarwal, NT. (2015). "MIND diet associated with reduced incidence of Alzheimer's disease". *Alzheimer's & dementia : the journal of the Alzheimer's Association* 11.9, pp. 1007–1014.

Nasserinejad, K., Rosmalen, J. van, Kort, W. de, and Lesaffre, E. (2017). "Comparison of criteria for choosing the number of classes in Bayesian finite mixture models". *PLoS One* 12.1. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0168838.

National_Institute_on_Aging (2017). "How Is Alzheimer's Disease Diagnosed?" https://www.nia.nih.gov/health/how-alzheimers-disease-diagnosed.

Papageorgiou, G., Mauff, K., Tomer, A., and Rizopoulos, D. (2019). "An overview of joint modeling of time-to-event and longitudinal outcomes". *Annual review of statistics and its application* 6.1, pp. 223–240.

Petersen, RC., Aisen, PS., Beckett, LA., Donohue, MC., Gamst, AC., Harvey, DJ., Jack, CR., Jagust, WJ., Shaw, LM., Toga, AW., Trojanowski, JQ., and Weiner, MW. (2010). "Alzheimer's disease neuroimaging initiative". *Neurology* 74.3, pp. 201–209.

Petersen, RC., Doody, R., and AKurz, A. (2001). "Current concepts in mild cognitive impairment". *Archives of neurology* 58.4, pp. 1985–1992.

Pinheiro, J. and Bates, D. (1995). "Approximations to the log-likelihood function in the nonlinear mixed-effects model". *Journal of computational and graphical statistics* 4, pp. 12–35.

Press, D. and Alexander, M. (2019). "Prevention of dementia". *Uptodate*.

Proust-Lima, C., Dartigues, J., and Jacqmin-Gadda, H. (2011). "Misuse of the linear mixed model when evaluating risk factors of cognitive decline". *American journal of epidemiology* 174.9, pp. 1077–1088.

Proust-Lima, C., Dartiguesa, J., and Jacqmin-Gadda, H. (2014). "Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: A latent process and latent class approach". *Statistics in medicine* 35, pp. 382–398.

Proust-Lima, C., Joly, P., Dartigues, JF., and Jacqmin-Gadda, H. (2009). "Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach". *Computational statistics and data analysis* 53.4, pp. 1142–1154.

Proust-Lima, C., Philipps, V., and Liquet, B. (2017). "Estimation of extended mixed models using latent classes and latent processes: The R package lcmm". *Journal of statistical software* 78.

Proust-Lima, C., Sene, M., Taylor, JM., and Jacqmin-Gadda, H. (2014). "Joint latent class models for longitudinal and time-to-event data: A review". *Statistical methods in medical research* 23.1, pp. 74–90.

Raboud, J., Reid, N., Coates, RA., and Farewell, VT. (1993). "Estimating risks of progressing to AIDS when covariates are measured with error". *Journal of the royal statistical society* 156.3, pp. 393–406.

Rizopoulos, D. (2010). "JM: An R Package for the joint modelling of longitudinal and time-to-event data". *Journal of statistical software* 35.9, pp. 1–33.

— (2014). "The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC". *Journal of statistical software* 72, pp. 1–45.

Rizopoulos, D. and Ghosh, P. (2011). "A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event". *Statistics in medicine* 30.12, pp. 1366–1380.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). "Shared parameter models under random effects misspecification". *Biometrika* 95.1, pp. 63–74.

Rouanet, A., Joly, P., Dartigues, J., Proust-Lima, C., and Jacqmin-Gadda, H. (2016). "Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia". *Biometrics* 72.4, pp. 1123–1135.

Sabia, S., Dugravot, A., Dartigues, J., Abell, J., Elbaz, A., Kivimäki, M., and Singh-Manoux, A. (2017). "Physical activity, cognitive decline and risk of dementia: 28-year follow-up of Whitehall II cohort study". *British medical journal* 357, J2709.

Salthouse, TA. (2009). "When does age-related cognitive decline begin?" *Neurobiology of aging* 30.4, pp. 507–514.

Satagopan, JM., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., and Auerbach, AD. (2004). "A note on competing risks in survival data analysis". *British journal of cancer* 91, pp. 1229–1235.

Segalas, C., Amieva, H., and Jacqmin-Gadda, H. (2019). "A hypothesis testing procedure for random changepoint mixed models". *Statistics in medicine* 38.20, pp. 3791–3803.

Segalas, C., Helmer, C., and Jacqmin-Gadda, H. (2020). "A curvilinear bivariate random changepoint model to assess temporal order of markers". *Statistical methods in medical research*.

Self, S. and Pawitan, Y. (1992). "Modeling a marker of disease progression and onset of disease". *AIDS epidemiology* 156, pp. 231–255.

Sliwinski, MJ., Stawski, RS., Hall, CB., Katz, M., Verghese, J., and Lipton, R. (2006). "Distinguishing preterminal and terminal cognitive decline". *European psychologist* 11.3, pp. 172–181.

Smith, A. (1975). "A Bayesian approach to inference about a change-point in a sequence of random variables". *Biometrika* 62.2, pp. 407–416.

Stephens, M. (2000). "Dealing with label switching in mixture models". *Journal of the royal statistical society, Series B (Statistical Methodology)* 62.4, pp. 795–809.

Stong, C. (2013). "Are some people immune to dementia?" *Neurology reviews* 21.9, p. 13.

Sturtz, S., Ligges, U., and Gelman, A. (2005). "R2winbugs: A package for running winbugs from r". *Journal of statistical software* 12, pp. 1–16.

Tan, C., Yu, J., Wang, H., Tan, M., Meng, X., Wang, C., Jiang, T., Zhu, X., and Tan, L. (2014). "Efficacy and safety of donepezil, galantamine, rivastigmine, and memantine for the treatment of Alzheimer's disease: a systematic review and meta-analysis". *Journal of Alzheimer's disease* 41.2, pp. 615–631.

Tsiatis, A. and Davidian, M. (2004). "Joint modeling of longitudinal and time-to-event Data: An Overview". *Statistica sinica* 14, pp. 809–834.

Tsiatis, AA., DeGruttola, V., and Wulfsohn, MS. (1995). "Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS". *Journal of the American statistical association* 90.429, pp. 27–37.

Watanabe, S. (2010). "Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory". *Journal of machine learning research* 11, pp. 3571–3594.

White, A. and Murphy, T. (2014). "BayesLCA: An R package for Bayesian latent class analysis". *Journal of statistical software* 61, pp. 1–28.

Wilson, RS., Beckett, LA., Bienias, JL., Evans, DA., and Bennett, DA. (2003). "Terminal decline in cognitive function". *Neurology* 60.11, pp. 1782–1787.

Wilson, RS., Segawa, E., Boyle, PA., Anagnos, SE., Hizel, LP., and Bennett, DA. (2013). "The natural history of cognitive decline in Alzheimer's disease". *Psychology and aging* 27.4, pp. 1008–1017.

Wilson, RS., Wilson, RS., Beck, TL., Bienias, JL., and Bennett, DA. (2007). "Terminal cognitive decline: accelerated loss of cognition in the last years of life". *Psychosomatic medicine* 69.2, pp. 131–137.

Wilson, RS., Yu, L., Lamar, M., Schneider, JA., Boyle, PA., and Bennett, DA. (2019). "Education and cognitive reserve in old age". *Neurology* 92.10, e1041–e1050.

Wulfsohn, MS. and Tsiatis, AA. (1997). "A joint model for survival and longitudinal data measured with error". *Biometrics* 53.1, pp. 330–339.

Wyse, J. and Friel, N. (2012). "Block clustering with collapsed latent block models". *Statistics and computing* 22, pp. 415–428.

Xu, D., Rhee, J., Mehta, RK., and Srinivasan, D. (2018). "Neuromuscular control and performance differences associated with gender and obesity in fatiguing tasks performed by older adults". *Frontiers in physiology* 9.800.

Yang, L. and Gao, S. (2013). "Bivariate random change point models for longitudinal outcomes". *Statistics in medicine* 32.6, pp. 1038–1053.

Yang, L., Yu, M., and Gao, S. (2016). "Joint models for multiple longitudinal processes and time-to-event outcome". *Journal of statistical computation and simulation* 86.18, pp. 3682–3700.

Yin, G. and Ibrahim, GJ. (2009). "Cure rate models: A unified approach". *Statistical society of Canada* 33, pp. 559–570.

Yu, B. and Ghosh, P. (2010). "Joint modeling for cognitive trajectory and risk of dementia in the presence of death". *Biometrics* 66.1, pp. 294–300.

Yu, L., Boyle, P., Schneider, JA., Segawa, E., Wilson, RS., Leurgans, S., and Bennett, DA. (2013). "Alzheimer's disease pathology, cerebrovascular disease, and cognitive change over the years prior to death". *Psychology and aging* 28.4, pp. 1015–1023.

Zhang, R., Simon, G., and Yu, F. (2017). "Advancing Alzheimer's research: a Review of big data promises". *International journal of medical informatics* 106, pp. 48–56.

Zhou, XA. (2013). "Modeling time to Alzheimer's disease with an immune model". `https://files.alz.washington.edu/presentations/2013/fall/Zhou.pdf`.

APPENDICES

Appendix A

Cure Rate Model

Cure rate models are motivated by relatively common practical example where a fraction of the subjects continue to survive from the event even after a long follow-up. The "cure rate" is defined as an asymptotic value of the survival function and is not directly observed. Cure rate model has been applied in melanoma and breast cancer research (Cooner et al., 2018). Zhou, 2013, introduced the concept of immune subgroup for AD research and modeled time to AD with an immune model, which illustrated that standard survival models may give misleading results. To apply the cure rate survival model to our case, we used a mixture cure model (MCM), which has been widely applied for survival analysis with a cure fraction (Berkson and Gage 1952). This model postulates that an immune fraction, c, of the subjects are cured, immune, or failure-free from the risk (event=0 forever). The survival time of the non-immune subjects follows a proper distribution with survival function $S_0(t)$. The two groups are modeled jointly in a MCM as:

$$S_{all}(t \mid z) = c + (1 - c)S_0(t \mid z). \tag{A.6}$$

As the first step, we apply the MCM on the analysis of time to AD. Following of work of Yin and Ibrahim, 2009, who developed a general class of cure models through transformation on the population COX survival function, we have likelihood as:

$$\begin{aligned} &L_1 = -\theta[1 - S(t)] \; (for\ all\ observations) \\ &L_2 = [x\beta + log\rho + (\rho - 1)log(t) + \eta - e^{\eta} * t^{\rho}] * Event^{i} \; (for\ events) \\ &where,\ x\beta = \alpha_0 + \alpha_1 * age + \alpha_2 * gender + \alpha_3 * educ \\ &\theta = e^{x\beta} \; (link\ function) \end{aligned} \tag{A.7}$$

We use Bayesian framework with first-activation scheme setting assuming a single activation leads to observed failure. Table XXIII provides the posterior means of the parameters. Unexpect-

Appendix A (Continued)

edly, we estimate a rather small fraction for c, which is less than 0.01. We next examine the survival probability plot for our data and we see that the survival probability across years is decreasing at a constant rate and there is no trend of slowing down at later time (Figure. 16a). Cure rate models are often used to fit the data that show a plateau (Figure 16b red) of the estimated survival curve at a value strictly greater than zero. When the plateau of the survival curve is not observed, the standard survival model may have an acceptable fit before censoring occurs. Based on this result, we decide not to incorporate a cure fraction on our joint modeling analysis.

Table XXIII: CURE RATE FRACTION MODEL FOR AD INCIDENCE

Parameter	Estimate	95% CI
$Intercept$	0.852	(-3.128, 2.336)
$Agefb$	0.107	(0.089, 0.125)
$Gender$	-0.003	(-0.294, -0.271)
$Educ$	-0.002	(0.038, 0.036)
η	-11.650	(-14.130, -9.611)
ρ	1.850	(1.680, 2.031)
θ	1722	
$var(\theta)$	1387	
$Fraction(c)$	<0.01	

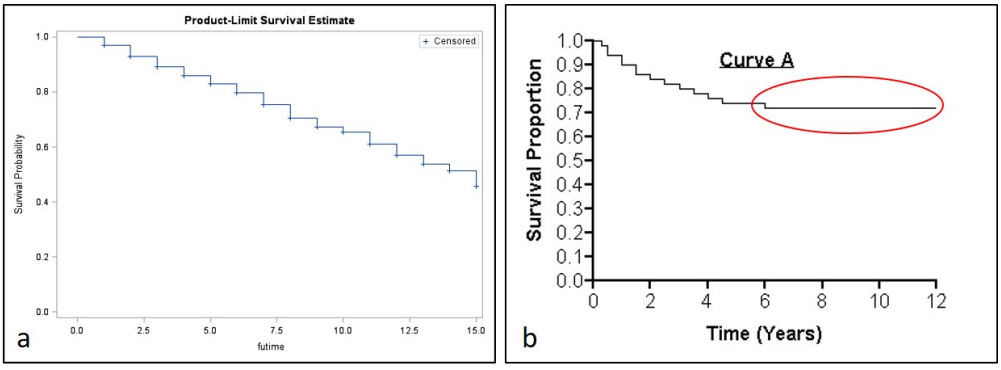

Figure 16: Comparison of survival curves
a. MAP cohort b. Example of plateau in survival curve

Appendix B

Trace and Density Plots of Markov Chain for Selected Parameters in Chapter 3

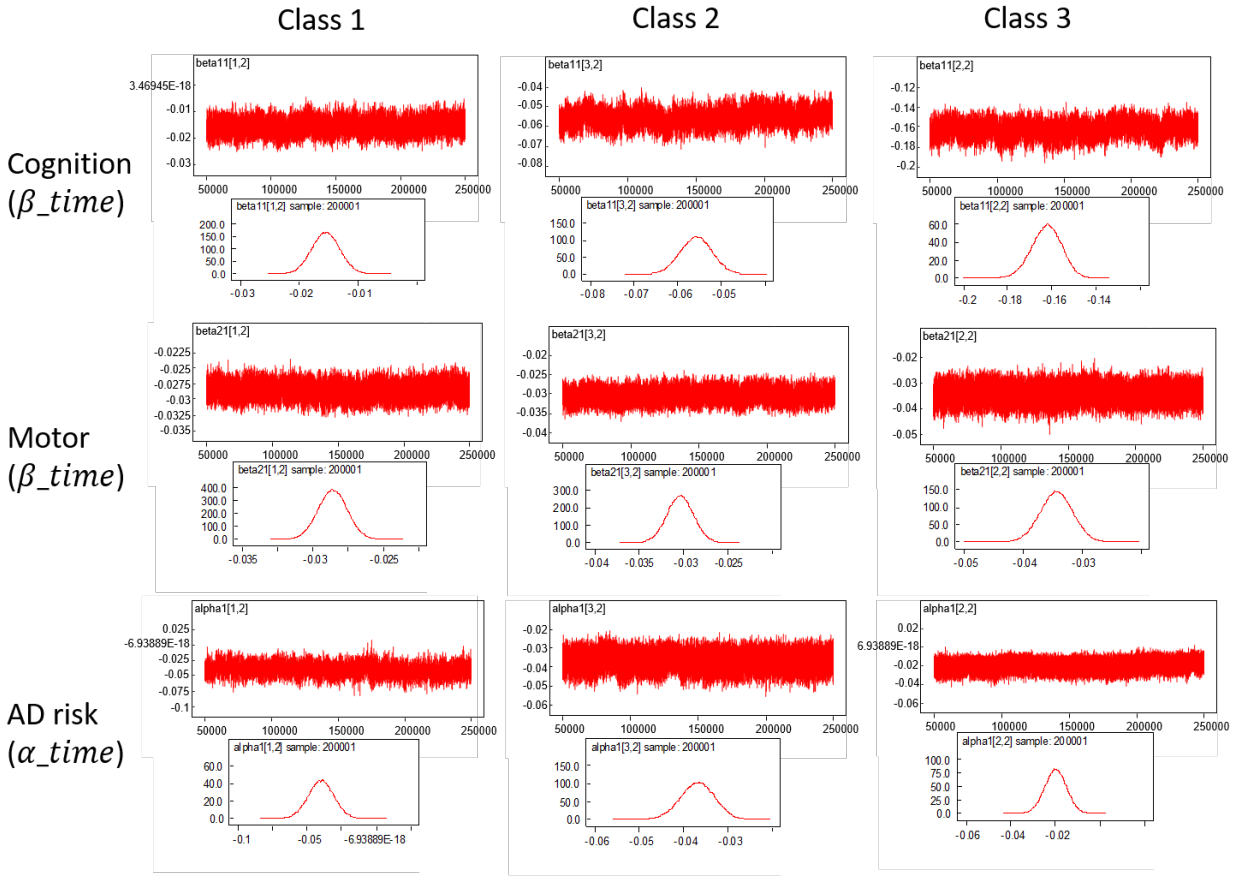

Figure 17: Trace and density plots of Markov chains for selected parameters for model of cognition and motor

Appendix B (Continued)

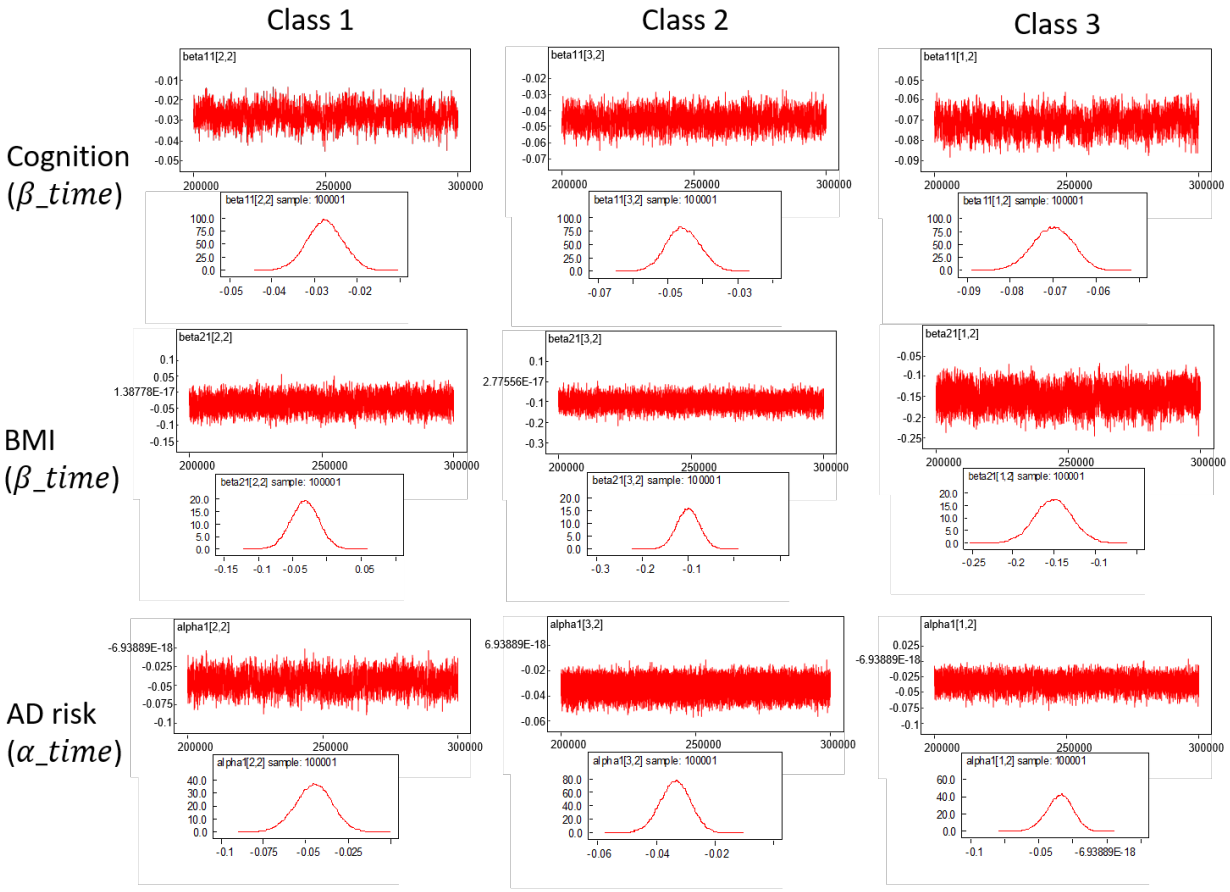

Figure 18: Trace and density plots of Markov chains for selected parameters for model of cognition and BMI

Appendix C

Comparison of the Baseline Characteristics among Latent Classes in Chap 3

Table XXIV: COMPARISON OF THE BASELINE CHARACTERISTICS AMONG CLASSES

Variable	Class 1	Class 2	Class 3	P-value
Model of cognition and motor				
Age_bl, mean	77.18	78.42	81.29	0.01
Male, %	22.48	25.00	19.61	0.55
Educ, mean	15.09	15.06	14.69	0.49
ApoE4, %	17.62	23.25	27.18	0.05
Model of cognition and BMI				
Age_bl, mean	77.24	78.07	79.03	0.03
Male, %	22.48	25.00	19.61	0.55
Educ, mean	15.19	14.91	14.92	0.51
ApoE4, %	25.52	20.13	21.61	0.36

Appendix D

Simulation Results for Scenario 3 in Chapter 4

Table XXV: SIMULATION STUDY RESULT FOR JOINT MODEL WITH BENT-CABLE FUNCTION (SCENARIO 3B, $\gamma \sim U(0, 10)$)

Parameter	True	Post mean	SD	95%CI
Longitudinal				
Outcome1				
β_{10}	1.0	1.00	0.09	(0.93)
β_{11}	-0.1	-0.10	0.00	(0.94)
β_{12}	-0.5	-0.50	0.01	(0.90)
γ_1	2.0	2.01	0.10	(0.95)
$\sigma_{\epsilon 1}$	0.2	0.20	0.002	(0.95)
Outcome2				
β_{20}	2	1.99	0.30	(0.91)
β_{21}	-0.2	-0.20	0.00	(0.91)
β_{22}	-0.6	-0.60	0.00	(0.92)
γ_2	1.0	0.89	0.27	(0.92)
$\sigma_{\epsilon 2}$	0.5	0.50	0.005	(0.90)
Cov_{10_20}	0.010	0.0101	0.0033	(0.90)
Cov_{10_21}	-0.001	-0.0009	0.0007	(0.93)
Cov_{11_20}	0.004	0.0043	0.0031	(0.88)
Cov_{11_21}	0.001	0.0009	0.0003	(0.86)
Changepoint1				
$\beta_{\tau_1 0}$	90.0	90.00	0.17	(0.94)
$\beta_{\tau_1 1}$	2.0	2.02	0.24	(0.95)
σ_{τ_1}	3	3.02	0.12	(0.95)
Changepoint2				
$\beta_{\tau_2 0}$	86.0	86.00	0.23	(0.96)
$\beta_{\tau_2 1}$	-2.0	-1.99	0.23	(0.97)
σ_{τ_2}	4	4.02	0.15	(0.95)
ρ	0.83	0.83	0.02	(0.95)
Survival				
Event1				
ζ_1	-0.5	-0.52	0.07	(0.93)
Event2				
ζ_2	-1.0	-1.01	0.06	(0.89)

Appendix E

Trace and Density Plots of Markov Chain for Selected Parameters in Chapter 5

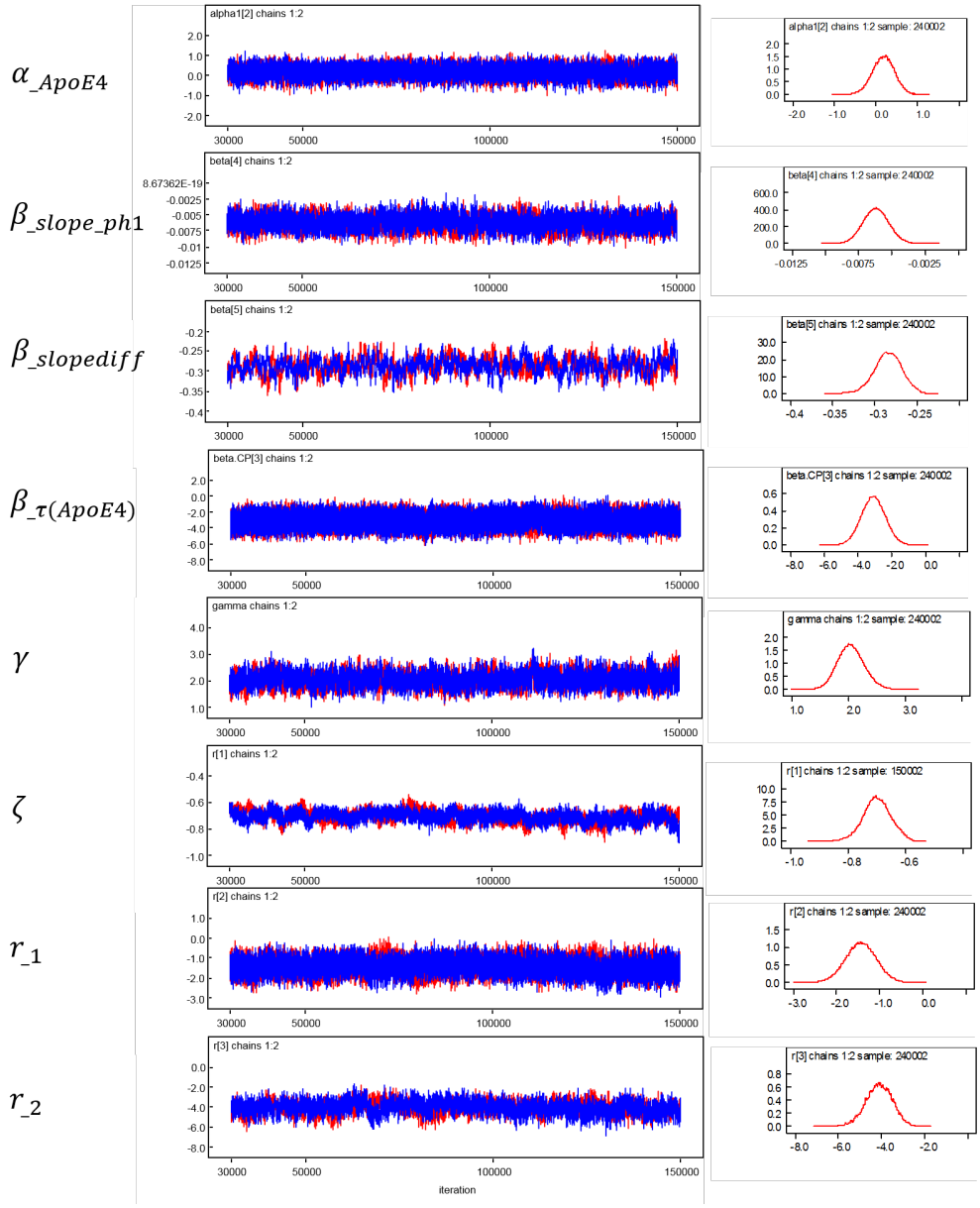

Figure 19: Trace and density plots of two Markov chains for selected parameters

Appendix F

WinBUGS Model for Joint Latent Class Model

```
# Each ID [i] belongs to one or several of K latent classes
 for (i in 1:N) {

# K Latent class classification
        class[i] ~ dcat(p[i,1:K])

      # Class membership probability
        for (k in 1:K) {
        p[i,k] <- xi[i,k]/sum(xi[i,])
        log(xi[i,k])<-int[k] + slp[k]*cov1[i]
                        } #end for k
      # Same as p[1:K] ~ ddirch(pi[]) if no covariate

# Class-specific regression coefficients in longitudinal model
       Beta11[i,1] <- beta11[class[i],1]  # Intercept
       Beta11[i,2] <- beta11[class[i],2]  # Time
       Beta21[i,1] <- beta21[class[i],1]  # Intercept
       Beta21[i,2] <- beta21[class[i],2]  # Time

# Class-specific alpha coefficients in survival model
       Alpha1[i,1] <- alpha1[class[i],1]  # intercept
       Alpha1[i,2] <- alpha1[class[i],2]  # Time

# Longitudinal model for continuous outcomes Y1 and Y2
   for (j in 1:M)  {
```

Appendix F (Continued)

```
    mu1[i,j] <- Beta11[i,1] + Beta11[i,2]*t[j] + beta1*cov2[i]
                + b1[i,1] + b1[i,2]*t[j]
    Y1[i,j] ~ dnorm(mu1[i,j],tau.Y1)

    mu2[i,j] <- Beta21[i,1] + Beta21[i,2]*t[j] + beta2*cov3[i]
                + b2[i,1] + b2[i,2]*t[j]
    Y2[i,j] ~ dnorm(mu2[i,j], tau.Y2)
                    } #end for j

# Weibull survival model
    Tevent[i] ~ dweib(r, mu.eta[i])I(Tcensor[i],)
    mu.eta[i] <- pow(mu.etachg[i], -r)
    mu.etachg[i] <- exp(Alpha1[i,1] + Alpha1[i,2]*cov4[i]
                    + alpha2[1]*cov5[i])

# Normal random effects
    b1[i,1:2] ~ dmnorm(mu.b1[],tau.rand1[,])
    b2[i,1:2] ~ dmnorm(mu.b2[],tau.rand2[,])

# Class proportion
    N2[i] <- (class[i]-1)*(3-class[i])
    N3[i] <- (class[i]-1)*(class[i]-2)/2
            } #end for i
    P2 <- (sum(N2[]))/N
    P3 <- (sum(N3[]))/N
    P1 <- 1-P2-P3
```

Appendix G

WinBUGS Model for Bent-Cable Regression

```
# Joint model with changepoint using bent-cable function
# Age is used as the time scale
  for (i in 1:N){
   for (j in 1:M) {

# Longitudinal model
   Y[i,j] ~ dnorm(mucog[i,j], tau.Y)

   x0[i,j] <- (time[i,j]-CP[i]+gamma)
   x1[i,j] <- pow(x0[i,j],2)/(4*gamma)
              *(1-step(CP[i]-gamma-time[i,j]))
              *(1-step(time[i,j]-CP[i]-gamma))
   x2[i,j] <- (time[i,j]-CP[i])*step(time[i,j]-CP[i]-gamma)
   x[i,j] <- x1[i,j] + x2[i,j]

   Y[i,j] <- beta[1] + beta[2]*cov1[i] + Rand[i,1]\\
             + beta[3]*time[i,j] + Rand[i,2]\\
             + beta[5]*x[i,j] + Rand[i,4]*x[i,j]
                    } #end of j

# Regression of CP
   ageCP[i] ~ dnorm(meanCP[i], tau.CP) I(min, max)
   meanCP[i] <- beta.CP[1] + beta.CP[2]*cov2[i]

# Random effects
```

Appendix G (Continued)

```
    Rand[i,1:3] ~ dmnorm(mu.rand[], tau.rand[,])


# Survival model (competing risk and interval censor)
# Transformed CP
    mux[i,1]<-exp(alpha1[1]*cov3[i]
                + r[1]*CP[i] + r[2]*Rand[i,1]+ r[3]*Rand[i,2])
    mux[i,2]<-exp(alpha2[1]*cov4[i]
                + r[4]*CP[i] + r[5]*Rand[i,1]+ r[6]*Rand[i,2])
    for (k in 1:2){
    base[i,k] <- lambda[k]*alpha[k]*pow(t2[i], alpha[k]-1)
    h[i,k ] <- base[i,k]*mux[i,k]
                    } #end of k


    logsurv[i,1] <- log(exp(-lambda[1]*pow(t1[i], alpha[1])*mux[i,1])
                      -exp(-lambda[1]*pow(t2[i], alpha[1])*mux[i,1]))
    logsurv[i,2] <- -lambda[2]*pow(t2[i], alpha[2])*mux[i,2]


    logL[i] <- delta[i,2]*log(h[i,2]) + delta[i,1]*logsurv[i,1]
                + (delta[i,1]-1)*(delta[i,2]-1)*logsurv[i,2]
                + delta[i,2]*logsurv[i,2]


    phi[i] <- 1000 - logL[i]
    zero[i] ~ dpois(phi[i])
         } #end of i
```

VITA

NAME: Yamin Wang

EDUCATION: MS, Biostatistics,
University of Illinois at Chicago,
Chicago, IL, 2013

PhD, Biochemistry and Molecular Biology,
Rosalind Franklin University of Medicine and Science,
North Chicago, IL, 2008

WORKING: Rush University Medical Center, Chicago, IL

Statistician III (August 2018 – present),

Statistician II (August 2013 – August 2018)