Journal of Statistical Research *https://doi.org/10.3329/jsr.v58i1.75413*
2024, Vol. 58, No. 1, pp. 49-73 *ISSN 0256 - 422 X*

JOINT MODELING OF VARYING-DISEASE-STATE LONGITUDINAL ORDINAL DATA AND TIME-TO-EVENT DATA WITH APPLICATION TO ALZHEIMER'S DISEASE

XIAOLIN CHANG

Department of Statistics, University of Connecticut, Storrs, Connecticut, U.S.A.

Email: xiaolin.chang@uconn.edu

MING-HUI CHEN*

Department of Statistics, University of Connecticut, Storrs, Connecticut, U.S.A.

Email: ming-hui.chen@uconn.edu

FOR THE ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE[1]

SUMMARY

Joint models are routinely used in clinical trials to fit longitudinal and survival data simultaneously in an integrated fashion. We propose a joint model that incorporates state-specific trajectories for fitting the longitudinal ordinal response and time-to-event data, focusing on its application to Alzheimer's Disease (AD). The proposed joint model effectively captures the fluctuating cognitive conditions observed before and after the transition between two disease states. By integrating longitudinal data into the survival sub-model through shared trajectories, we can improve the fit of the survival data. A Markov chain Monte Carlo (MCMC) sampling algorithm is developed to carry out Bayesian computation. A variation of the Deviance Information Criterion is developed to assess the fit of each component of the joint model as well as the contribution of the longitudinal data in fitting the survival data. A variation of the concordance (C) index is further derived to assess the discriminatory and predictive performance of the longitudinal sub-model. An in-depth analysis of the real data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database is carried out to demonstrate the applicability of the proposed methodology.

Keywords and phrases: Disease progression; Joint model; Longitudinal ordinal data; State-specific trajectory; Time-to-event data; Transition of state.

* Corresponding author

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

1 Introduction

Medical studies often collect longitudinal biomarker measurements and time-to-event data. Typically, these types of data are analyzed separately using conventional methods such as the linear mixed effects models (Laird and Ware, 1982) for longitudinal data (Guerrero et al., 2016; Bernal-Rusiel et al., 2013b) and Cox proportional hazards models (Cox, 1972) for survival data. However, modeling them separately can lead to biased parameter estimates (Hu et al., 2009) due to measurement errors and the failure to account for the endogeneity of the longitudinal evolution. To address these limitations, joint modeling (Chen et al., 2004; Ibrahim et al., 2004; Tsiatis and Davidian, 2004) has emerged as a popular approach in many medical studies. Joint models integrate the longitudinal and survival models by incorporating association parameters and fitting them simultaneously. This approach allows for the assessment of the association between the longitudinal and survival outcomes, leading to more accurate estimations. By considering the longitudinal measurements in the survival model, the overall fit is improved. Additionally, the contribution of the longitudinal data in the fit of the survival data can be quantified using DIC factorization.

The literature on joint models encompasses a wide range of medical studies and numerous extensions. An overview of the joint models is provided in Papageorgiou et al. (2019) and Tsiatis and Davidian (2004). Sheikh et al. (2022) introduced a power prior approach for leveraging external data in joint models. Chen et al. (2004) proposed joint modeling of survival data with a cure fraction, while Wu et al. (2020), Li et al. (2019), Li et al. (2018a), Li et al. (2018b), and Ibrahim et al. (2004) applied joint models to various medical areas including Alzheimer's Disease. Joint models of multiple longitudinal biomarkers and survival data are introduced in Li et al. (2021) and Medina-Olivares et al. (2023). Huang et al. (2011), Sheikh et al. (2021) and Zhang et al. (2021) developed joint models of survival data with competing or semi-competing risks. A joint model of functional data and survival data was presented in Li et al. (2022). Andrinopoulou et al. (2021) and Li et al. (2017) described predictions using joint models. Zhang et al. (2017) and Zhang et al. (2014) presented model assessments, while Zhang et al. (2016) and Sheikh et al. (2021) focused on computational methods and software development.

While the conventional linear mixed effects model is commonly used for modeling longitudinal outcomes (Yi and Tang, 2022; Ziegler et al., 2015; Bernal-Rusiel et al., 2013a), it may not capture the fluctuations of biomarkers before and after state transitions in diseases involving such transitions. To address this limitation, our proposed method incorporates state-specific fixed and random intercepts and slopes that consider the shift in longitudinal measurements during state transitions. Instead of using real time, we utilize the cumulative time that a participant stays in the current state, resetting it to zero after a state switch to start a new trajectory. The survival sub-model is implemented using the Cox proportional hazards model. We choose the shared trajectory (Ibrahim et al., 2004) as the association term that links the longitudinal and survival sub-models, as it provides a more straightforward and meaningful interpretation.

Ordinal longitudinal data are commonly encountered in a medical study, given the inherent order that ranges from the lowest to the highest level. Numerous methods have been proposed for modeling ordinal data in the existing literature. Some classic models are introduced in McCullagh (1980). Various Bayesian approaches are presented in Albert and Chib (1993), Cowles et al. (1996) and

Chen et al. (2000). Furthermore, the literature exploring the modeling of longitudinal ordinal data includes Molenberghs et al. (1997), Lee and Daniels (2008) and Varin and Czado (2010), while discussions about ordinal regression with mixed effects can be found in Hedeker and Gibbons (1994) and Liu and Hedeker (2006). The joint model of longitudinal ordinal data and survival times is proposed in Li et al. (2010). Some applications in medical studies are presented in Lee and Daniels (2008) and Jacqmin-Gadda et al. (2010). In our proposed joint model, a proportional odds mixed-effects ordered probit sub-model with state-specific trajectories with is assumed for the longitudinal ordinal data.

For Bayesian computation, we develop a Markov chain Monte Carlo (MCMC) sampling algorithm (Chen et al., 2000). Due to the inability to directly sample most parameters with the shared trajectory, we employ a localized Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970). To assess the fit of each component of the joint model and the contribution of longitudinal data in fitting the survival data, we develop a variation of the Deviance Information Criterion (DIC). For the longitudinal response data, a variation of the concordance index is further derived to assess the discriminatory and predictive performance of the longitudinal sub-model.

The remainder of the paper is organized as follows. Section 2 provides a description of the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Section 3 describes the proposed joint model with state-specific trajectories for fitting longitudinal ordinal biomarker and time-to-event data. Section 4 presents the prior and posterior, the MCMC sampling algorithm, the DIC criteria and Bayesian concordance index. The empirical results of the analysis of the ADNI data are given in Section 5. We conclude the paper with brief discussion and future research in Section 6. Some computational details and proofs are given in the Appendix.

2 Motivating ADNI data

The following statements are quoted from `https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf`:

> "Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org."

The ADNI database comprises data from more than 2000 participants aged 55 or above. The dataset includes three primary diagnosed disease states: Cognitive Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD). Our research aims to investigate the association between longitudinal ordinal cognitive assessment and the progression to AD. We define the event as AD-free survival that the event indicator is set to 1 if the patient is diagnosed with AD or de-

ceased; and 0 if the patient is alive and free from AD. Since progression to AD is treated as the event, the participants diagnosed with AD at baseline are excluded. For participants transitioning from CN/MCI to AD, we exclude the longitudinal measurements taken after their AD diagnosis. Therefore, the potential transitions considered in our study include staying in CN, staying in MCI, transitioning from CN to MCI, and transitioning from MCI to CN. Figure 1 shows the frequency of each type of state transition either remaining in or transitioning between CN, MCI, and AD. We can see that majority of the transitions are remaining in the same state. There are more than 100 occurrences for the transition between CN and MCI. More than 300 transitions from MCI to AD are observed. Additionally, the number of participants following different patterns of disease

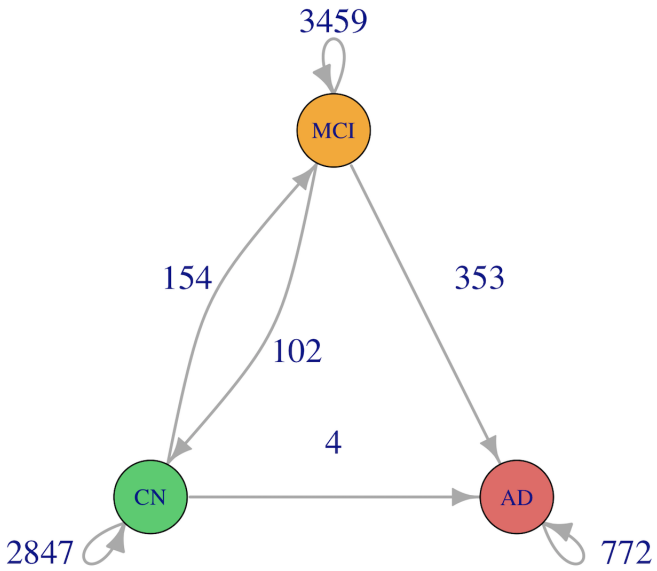

Figure 1: Frequency of each type of state transition either remaining in or transitioning between CN, MCI, and AD.

progression are summarized in Table 1.

The longitudinal ordinal measurement we are interested in studying is a cognitive assessment known as the Clinical Dementia Rating (CDR). The CDR rates five degrees of impairment in performance on each of the 6 categories of cognitive function including memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. The ratings of degree of impairment obtained on each of the 6 categories of function are synthesized into one global rating of dementia (ranged from 0 to 3 with higher value means worse cognitive condition). This is used as a global measure of severity of dementia. CDR originally has five ordinal categories observed in ADNI, namely 0 is Normal, 0.5 is Very Mild Dementia, 1 is Mild Dementia, 2 is Moderate Dementia, and 3 is Severe Dementia. Due to the setting of our study, after excluding the participants whose baseline diagnoses are AD, we only have the three least severe categories of CDR observed.

Table 1: Number of participants following different patterns of disease progression.

Pattern	Number of participants
Staying in CN	506
CN-MCI	100
CN-MCI-CN	7
CN-MCI-CN-MCI	8
CN-MCI-CN-MCI-CN	1
Staying in MCI	746
MCI-CN	50
MCI-CN-MCI	16
MCI-CN-MCI-CN	6
MCI-CN-MCI-CN-MCI	3

The number of participants who have CDR measured at each visit is shown is Figure 2. Most of the

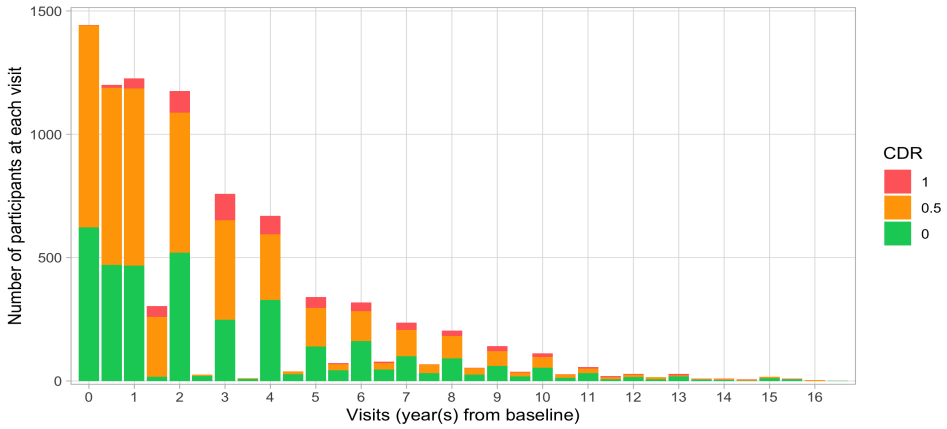

Figure 2: Number of participants who have CDR measured at each visit.

follow-up visits occur annually, although some participants have semi-annual visits based on their cognitive examination results. Given the longitudinal nature of CDR measurements, we employ a state-specific trajectory model as the longitudinal sub-model. For the survival data, the diagnosis of AD or death is considered as the event. Regarding the accuracy of the diagnosis time of AD, we examine the time to reach the Minimal Clinically Important Difference (MCID) for AD and the time gap between the AD diagnosis and the last prior visit in ADNI data. According to the literature, the time to reach the MCID for AD in terms of the CDR is less than one year. The longitudinal measure-

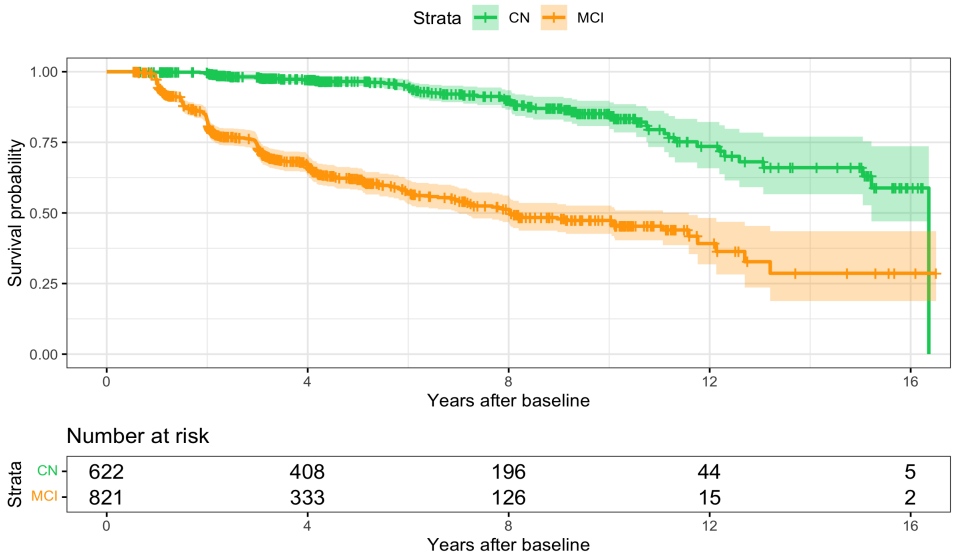

Figure 3: Kaplan-Meier curves for the event of progression to AD or death by diagnosed states at baseline.

ment used in this paper is the global CDR, which is a higher-level summary scale for CDR Sum of Boxes (CDR-SB). According to Cummings (2023), the MCID for CDR-SB is 1.19. The annual rate of change in CDR-SB is 1.43 (SE = 0.05) for the global CDR 0.5 cohort; and 1.91 (SE = 0.07) for the global CDR 1 cohort (Williams et al., 2013). For the time gap between the AD diagnosis and the last prior visit, the Q1, median, and Q3 of the time gaps in years are 0.41, 0.65, and 0.75, respectively. In terms of the visit code, 54% of the gap is a half-year visit and 90% is within a one-year visit. Given that the majority of the time gaps between the AD diagnosis and the last prior visit are within a year, and considering the time to reach MCID in terms of CDR is less than a year, it is reasonable to assume that the time of AD diagnosis is accurate. Figure 3 presents the Kaplan-Meier survival curves for AD-free survival by different diagnosed states at baseline. The participants diagnosed as MCI at baseline have a significantly higher probability of progressing to AD or death. For the fitting of time to AD, we use the Cox proportional hazard model as the survival sub-model. Then we can jointly model the longitudinal ordinal data and time-to-event data to investigate the association between the evolution of CDR and the progression to AD or death. Additionally, we extend the concept of concordance to ordinal responses for individual patient data. We consider seven baseline covariates in our analysis, including age (in years), sex (coded as 'Female' = 1, 'Male' = 0), race (coded as 'White' = 1, 'Other' = 0), marital status (coded as 'Married' = 1, 'Other' = 0), education (in years), apolipoprotein E (APOE4) count, and Rey's Auditory Verbal Learning Test (RAVLT) forgetting percentages. These baseline covariates are used in both the longitudinal and survival sub-models. Table 2 presents a demographic summary of the baseline covariates by baseline diagnostic group, including the sample mean (sample standard deviation (SD)) for each of the continuous variables (age,

Table 2: A summary of the baseline covariates categorized by diagnostic group.

Baseline covariates	Total	CN at baseline	MCI at baseline	P-value
Age	73.3 (7.01)	73.4 (6.14)	73.2 (7.60)	0.9630
Gender				<0.001
Female	666 (46.1%)	335 (53.9%)	331 (40.2%)	
Male	779 (53.9%)	287 (46.1%)	492 (59.8%)	
Race				0.1500
White	1341 (92.8%)	570 (91.6%)	771 (93.7%)	
Other	104 (7.2%)	52 (8.4%)	52 (6.3%)	
Education	16.2 (2.75)	16.5 (2.63)	16.0 (2.81)	<0.001
Marital				0.0013
Married	1071 (74.1%)	434 (69.8%)	637 (77.4%)	
Other	374 (25.9%)	188 (30.2%)	186 (22.6%)	
APOE4				<0.001
0	842 (58.3%)	431 (69.3%)	411 (49.9%)	
1	493 (34.1%)	174 (28.0%)	319 (38.8%)	
2	110 (7.6%)	17 (2.7%)	93 (11.3%)	
RAVLT-forgetting	48.6 (32.6)	34.7 (27.9)	59.0 (32.0)	<0.001

education, APOE4, and RAVLT) and frequency (n) (relative frequency %) for each level of each of the categorical variables (sex, race, and marital status). In Table 2, the P-values are obtained from the Mann-Whitney test for numeric variables and Fisher's exact test for categorical variables. We observe statistically significant differences in gender (P-value < 0.001), education years (P-value < 0.001), marital status (P-value $= 0.001$), APOE4 count (P-value < 0.001), and RAVLT forgetting percentages (P-value < 0.001) between the Cognitive Normal (CN) and Mild Cognitive Impairment (MCI) groups at baseline. Furthermore, compared to the CN group, participants in the MCI group tend to have a lower proportion of females, lower level of education, higher proportion of being married, higher APOE4 count, and higher percentages of RAVLT forgetting.

3 Method

3.1 Longitudinal sub-model

Considering the longitudinal and ordinal nature of the measurements, we employ a state-specific trajectory model (SSTM) with ordered probit link as the longitudinal sub-model. Suppose we have m_i longitudinal measurements observed for the i-th subject for $i = 1, \ldots, n$. Let a_{ij} denote the time in years of the j-th visit from the baseline for $j = 0, 1, \ldots, m_i$ and $i = 1, \ldots, n$. Denote $\boldsymbol{y}_i^*(a_{ij})$ and $\boldsymbol{y}_i(a_{ij})$ as the latent and observed longitudinal measurements for subject i at time a_{ij}. The diagnosed state of subject i at time a_{ij} is represented by $\kappa_i(a_{ij})$. Specifically, $\kappa_i(a_{ij})$ equals 0 if subject i is diagnosed as Cognitive Normal (CN) at time a_{ij} and 1 if diagnosed as Mild Cognitive Impairment (MCI). Let $t^*(a_{ij})$ in (3.2) denote the cumulative time that subject i remains in the current state up to time a_{ij}, which is defined as

$$t^*(a_{ij}) = \begin{cases} 0, & \text{for } j = 0, \\ a_{ij} - \max\limits_{0 \le j^* \le j-1} \{a_{ij^*} 1\{\kappa_i(a_{ij^*}) = 1 - \kappa_i(a_{ij})\}\}, & \text{for } j = 1, \ldots, m_i. \end{cases} \tag{3.1}$$

Then, the longitudinal model is defined as

$$y_i^*(a_{ij}) = \theta_{\kappa_i(a_{ij})0} + \theta_{\kappa_i(a_{ij})0i} + (\theta_{\kappa_i(a_{ij})1} + \theta_{\kappa_i(a_{ij})1i}) t^*(a_{ij}) + \boldsymbol{x}_i' \boldsymbol{\gamma} + \epsilon_{ij}, \tag{3.2}$$

where $\theta_{\kappa_i(a_{ij})0}$ and $\theta_{\kappa_i(a_{ij})1}$ are the fixed intercept and fixed slope, respectively, specific to the diagnosed state $\kappa_i(a_{ij})$, similarly, $\theta_{\kappa_i(a_{ij})0i}$ and $\theta_{\kappa_i(a_{ij})1i}$ are the random intercept and random slope specific to the subject when diagnosed as state $\kappa_i(a_{ij})$, and $\boldsymbol{\gamma}$ is a p-dimensional vector of coefficients corresponding to the p-dimensional baseline covariates $\boldsymbol{x}_i$. After excluding the participants who are diagnosed as AD at the baseline, the CDR of the remaining patients contains three levels, namely 0, 0.5, and 1. Therefore, we have known cutoffs to be 0 and 1, while the variance of the latent variable remains unknown (Chen et al., 2000). The specific conversion between the latent

variable and the observed variable is

$$\boldsymbol{y}_i(a_{ij}) = \begin{cases} 0, & \text{if } \boldsymbol{y}_i^*(a_{ij}) < 0, \\ 0.5, & \text{if } 0 < \boldsymbol{y}_i^*(a_{ij}) < 1, \\ 1, & \text{if } \boldsymbol{y}_i^*(a_{ij}) > 1, \end{cases} \tag{3.3}$$

where $\boldsymbol{y}_i^*(a_{ij})$ is a latent measure defined in (3.2). The error term ϵ_{ij} is assumed to follow $N(0, \sigma^2)$. The random effects and the error term are assumed to be independent. Let $\boldsymbol{\theta}_i = (\theta_{00i}, \theta_{01i}, \theta_{10i}, \theta_{11i})' \sim N(\mathbf{0}, \Omega)$, where Ω is a positive-definite variance-covariance matrix. Using Cholesky decomposition of $\Omega = \Gamma\Gamma'$, we can reparameterize $\boldsymbol{\theta}_i$ (Sheikh et al., 2021) as follows

$$\boldsymbol{\theta}_i = \Gamma \boldsymbol{\theta}_i^R, \tag{3.4}$$

where $\boldsymbol{\theta}_i^R \sim N(0, I_4)$ and

$$\Gamma = \begin{pmatrix} b_{11} & 0 & 0 & 0 \\ b_{21} & b_{22} & 0 & 0 \\ 0 & 0 & b_{33} & 0 \\ 0 & 0 & b_{43} & b_{44} \end{pmatrix},$$

with positive diagonal entries. Now we can rewrite our longitudinal model (3.2) as

$$\begin{aligned} y_i^*(a_{ij}) = \theta_{\kappa_i(a_{ij})0} + (\Gamma\boldsymbol{\theta}_i^R)_{2\kappa_i(a_{ij})+1} + \big(\theta_{\kappa_i(a_{ij})1} + (\Gamma\boldsymbol{\theta}_i^R)_{2\kappa_i(a_{ij})+2}\big)t^*(a_{ij}) \\ + \boldsymbol{x}_i'\boldsymbol{\gamma} + \epsilon_{ij}. \end{aligned} \tag{3.5}$$

Figure 4 provides a visual representation illustrating that the model with state-specific trajectories yields a better fit compared to the model with a state-independent trajectory. At the top, middle, and bottom of Figure 4, we present the observed CDR, fitted CDR from the linear mixed model with an interaction between state and time, and the fitted CDR from the state-specific trajectory model, respectively, for a selected participant. The colors of the dots represent the diagnosed state for this patient at each time point. We can see that SSTM performs better when the patient transitions to CN at approximately year 4 and year 8 from the baseline.

Please note that this aforementioned model formulation assumes that both the biomarker and diagnosis are assessed during the same visit for ease of mathematical representation. However, in some cases, patients have separate visits for biomarker measurement and diagnosis. This means that state transitions could occur between two biomarker measurements. Thus, we have also developed a model that accommodates separate visits for biomarker measurement and diagnosis. The details of this model can be found in the appendix.

3.2 Survival sub-model

Let t_i be the event time or the censoring time and let δ_i represent the event indicator such that δ_i takes a value of 0 for censored and 1 for the event for the i-th subject. The survival sub-model incorporates

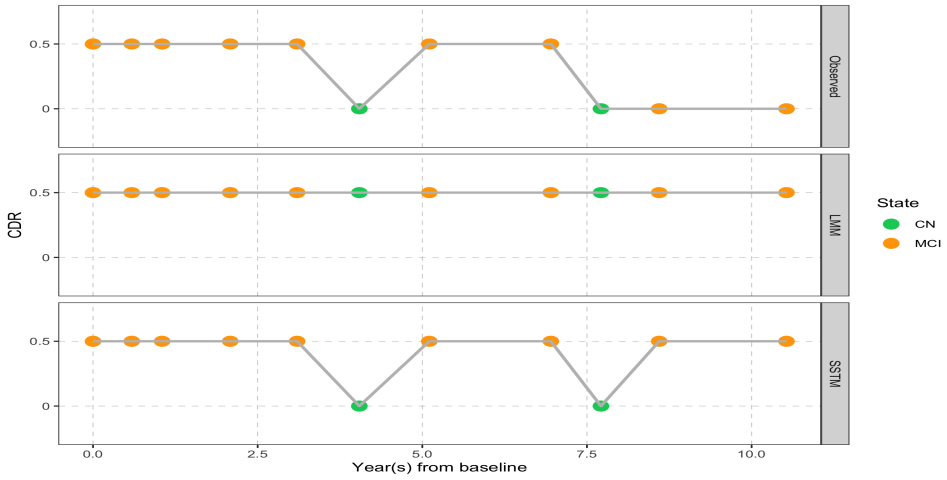

Figure 4: Trajectory of CDR from observed values (top), Linear Mixed Model with an interaction between state and time (middle), and State-Specific Trajectory Model (bottom).

the shared trajectory from the longitudinal sub-model, along with other fixed covariates. We assume a proportional hazards model (Cox (1972); Lin and Wei (1989)) with the hazard function given by

$$\mathrm{h}(t|\mathcal{X}_i(t), \boldsymbol{z}_i) = \mathrm{h}_0(t) \exp\Big\{\alpha_{\kappa_i(t)}\mathcal{X}_i(t) + \boldsymbol{z}_i'\boldsymbol{\beta}\Big\}, \tag{3.6}$$

where $\mathrm{h}_0(t)$ is the baseline hazard function, $\boldsymbol{\beta}$ is the vector of the coefficients corresponding to the baseline covariates $\boldsymbol{z}_i$, and $\alpha_{\kappa_i(t)}$ is the state-specific association parameter for the shared trajectory. In (3.6), α_0 and α_1 are the two association parameters corresponding to the states CN and MCI, respectively. Here, $\mathcal{X}_i(t)$ denotes the longitudinal trajectory at time t for subject i from the longitudinal model (3.2) and can be written as

$$\mathcal{X}_i(t) = \theta_{\kappa_i(t)0} + \theta_{\kappa_i(t)0i} + (\theta_{\kappa_i(t)1} + \theta_{\kappa_i(t)1i})t_i^*(t) + \boldsymbol{x}_i'\boldsymbol{\gamma}, \tag{3.7}$$

where $\kappa_i(t)$ is a function of time t and is defined the same way as in the longitudinal sub-model (3.2), and $t_i^*(t)$ is the cumulative time that the i-th subject remains in the current state until t, expressed as

$$t_i^*(t) = \begin{cases} 0, & \text{for } t = 0, \\ t - \max_{a_{ij^*} \le t}\{a_{ij^*} 1\{\kappa_i(a_{ij^*}) = 1 - \kappa_i(t)\}\}, & \text{for } t > 0. \end{cases} \tag{3.8}$$

Now the longitudinal model can be rewritten by $y_i^*(a_{ij}) = \mathcal{X}_i(a_{ij}) + \epsilon_{ij}$, as the only difference between $y_i^*(a_{ij})$ and $\mathcal{X}_i(a_{ij})$ is the error term. In the survival sub-model, the trajectory of the longitudinal biomarker builds a connection between the longitudinal and survival sub-models via the association parameter $\alpha_{\kappa_i(t)}$. With the reparameterization of $\boldsymbol{\theta}_i = \Gamma\boldsymbol{\theta}_i^R$ in (3.4), the shared

trajectory in the survival sub-model can be reformulated as

$$\begin{aligned}\alpha\mathcal{X}_i(t) =&\alpha\big\{\theta_{\kappa_i(t)0}+\theta_{\kappa_i(t)0i}+(\theta_{\kappa_i(t)1}+\theta_{\kappa_i(t)1i})t_i^*(t)+\boldsymbol{x}_i'\boldsymbol{\gamma}\big\}\\ =&\alpha\big\{\theta_{\kappa_i(t)0}+(\Gamma\boldsymbol{\theta}_i^R)_{2\kappa_i(a_{ij})+1}+(\theta_{\kappa_i(t)1}+(\Gamma\boldsymbol{\theta}_i^R)_{2\kappa_i(a_{ij})+1})t_i^*(t)+\boldsymbol{x}_i'\boldsymbol{\gamma}\big\}\\ =&1\{\kappa_i(t)=0\}\alpha\big\{\theta_{00}+b_{11}\theta_{i1}^R+(\theta_{01}+b_{21}\theta_{i1}^R+b_{22}\theta_{i2}^R)t_i^*(t)+\boldsymbol{x}_i'\boldsymbol{\gamma}\big\}\\ &+1\{\kappa_i(t)=1\}\alpha\big\{\theta_{10}+b_{33}\theta_{i3}^R+(\theta_{11}+b_{43}\theta_{i3}^R+b_{44}\theta_{i4}^R)t_i^*(t)+\boldsymbol{x}_i'\boldsymbol{\gamma}\big\}. \end{aligned} \tag{3.9}$$

The term for the fixed intercept in (3.9) can be absorbed by the baseline hazard. Thus, (3.7) can be rewritten as

$$\mathcal{X}_i(t)=\theta_{\kappa_i(t)0i}+(\theta_{\kappa_i(t)1}+\theta_{\kappa_i(t)1i})t_i^*(t)+\boldsymbol{x}_i'\boldsymbol{\gamma}. \tag{3.10}$$

It is also assumed that the baseline hazard function takes on a piecewise constant form with G partitions of the time axis such that $h_0(t)=\lambda_g, t\in(s_{g-1},s_g], g=1,2,\ldots,G, 0=s_0<s_1<\cdots<s_G=\infty$. Write $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\ldots,\lambda_G)$ as the vector of piecewise constants for the baseline hazard function.

3.3 Likelihood construction

Let $\boldsymbol{\theta}_{\text{CN}}=(\boldsymbol{\theta}_{00},\boldsymbol{\theta}_{01})$ and $\boldsymbol{\theta}_{\text{MCI}}=(\boldsymbol{\theta}_{10},\boldsymbol{\theta}_{11})$. Let $\boldsymbol{\phi}=(\boldsymbol{\phi}_1,\boldsymbol{\phi}_2)$, where $\boldsymbol{\phi}_1=(\boldsymbol{\theta}_{\text{CN}},\boldsymbol{\theta}_{\text{MCI}},\boldsymbol{\gamma},\sigma^2,\Gamma)$ and $\boldsymbol{\phi}_2=(\boldsymbol{\lambda},\boldsymbol{\alpha},\boldsymbol{\beta})$. For the longitudinal sub-model, the density of $\boldsymbol{y}_i^*$ conditional on $\boldsymbol{\theta}_i^R$ takes the following expression:

$$\begin{aligned}f(\boldsymbol{y}_i^*|\boldsymbol{\phi}_1,\boldsymbol{\theta}_i^R,\boldsymbol{\kappa}_i,\boldsymbol{x}_i)=\prod_{j=1}^{m_i}\prod_{k=0}^{1}(2\pi\sigma^2)^{-\frac{1\{\kappa_i(a_{ij})=k\}}{2}}\exp\Bigg\{&-\frac{1\{\kappa_i(a_{ij})=k\}}{2\sigma^2}\times\\ &\Big[y_i^*(a_{ij})-\big\{(\theta_{k0}+(\boldsymbol{\Gamma\theta}_i^R)_{2k+1})+(\theta_{k1}+(\boldsymbol{\Gamma\theta}_i^R)_{2k+2})t^*(a_{ij})+\mathbf{x}_i'\boldsymbol{\gamma}\big\}\Big]'\\ &\Big[y_i^*(a_{ij})-\big\{(\theta_{k0}+(\boldsymbol{\Gamma\theta}_i^R)_{2k+1})+(\theta_{k1}+(\boldsymbol{\Gamma\theta}_i^R)_{2k+2})t^*(a_{ij})+\mathbf{x}_i'\boldsymbol{\gamma}\big\}\Big]\Bigg\},\end{aligned} \tag{3.11}$$

where $t^*(a_{ij})$ is defined in (3.1) and the density of $\boldsymbol{\theta}_i^R$ is given by

$$f(\boldsymbol{\theta}_i^R)=\frac{1}{(2\pi)^2}\exp\{-\frac{1}{2}\boldsymbol{\theta}_i^{R'}\boldsymbol{\theta}_i^R\}. \tag{3.12}$$

The likelihood function of the survival sub-model for the i-th subject can be written as

$$\begin{aligned}f(t_i|\boldsymbol{\phi}_1,\boldsymbol{\phi}_2,\boldsymbol{\theta}_i^R,\delta_i,\boldsymbol{\kappa}_i,\boldsymbol{z}_i)=&\Big[\mathrm{h}_0(t_i)\exp\Big\{\alpha\mathcal{X}_i(t_i)+\boldsymbol{z}_i'\boldsymbol{\beta}\Big\}\Big]^{\delta_i}\\ &\times\exp\Big\{-\int_0^{t_i}h_0(u)\exp\big\{\alpha\mathcal{X}_i(u)+\boldsymbol{z}_i'\boldsymbol{\beta}\big\}du\Big\},\end{aligned} \tag{3.13}$$

where

$$\mathcal{X}_i(u)=(\theta_{\kappa_i(u),0}+(\boldsymbol{\Gamma\theta}_i^R)_{2\kappa_i(u)+1})+(\theta_{\kappa_i(u),1}+(\boldsymbol{\Gamma\theta}_i^R)_{2\kappa_i(u)+2}))t_{ih}^*(u)+\mathbf{x}_i'\boldsymbol{\gamma}, \tag{3.14}$$

and

$$t^*_{ih}(u) = u - \theta_{ih^*}, h^* = \text{argmax}_h\{\theta_{ih}|\theta_{ih} \le u\} \text{ and } t_i > 0. \tag{3.15}$$

The joint distribution of $(\boldsymbol{y}^*_i, t_i, \boldsymbol{\theta}^R_i)$ is given by

$$\begin{aligned} f(\boldsymbol{y}^*_i, t_i, \boldsymbol{\theta}^R_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) =& f(\boldsymbol{y}^*_i|\boldsymbol{\phi}_1, \boldsymbol{\theta}^R_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i)\times \\ & f(t_i|\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}^R_i, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{z}_i) f(\boldsymbol{\theta}^R_i), \end{aligned} \tag{3.16}$$

where $f(\boldsymbol{y}^*_i|\boldsymbol{\phi}_1, \boldsymbol{\theta}^R_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i)$ is defined in (3.12) and the marginal distribution of $(\boldsymbol{y}^*_i, t_i)$ can be written as

$$f(\boldsymbol{y}^*_i, t_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) = \int f(\boldsymbol{y}^*_i, t_i, \boldsymbol{\theta}^R_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\, d\boldsymbol{\theta}^R_i. \tag{3.17}$$

Furthermore, we can write

$$f(\boldsymbol{y}^*_i, t_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i) = f(\boldsymbol{y}^*_i|\boldsymbol{\phi}_1, \boldsymbol{\kappa}_i, \boldsymbol{x}_i) \times f(t_i|\boldsymbol{y}^*_i, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{z}_i), \tag{3.18}$$

where $f(\boldsymbol{y}^*_i|\boldsymbol{\phi}_1, \boldsymbol{\kappa}_i, \boldsymbol{x}_i) = \int f(\boldsymbol{y}^*_i|\boldsymbol{\phi}_1, \boldsymbol{\theta}^R_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i) f(\boldsymbol{\theta}^R_i) d\boldsymbol{\theta}^R_i$ is the marginal distribution of $\boldsymbol{y}^*_i$ and $f(t_i|\boldsymbol{y}^*_i, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{z}_i)$ is the conditional distribution of t_i given $\boldsymbol{y}^*_i$. Let $\text{D}_{\text{obs}} = \{(\boldsymbol{y}_i, t_i, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i), i = 1, \ldots, n\}$ denote the observed data. Then the likelihood is given by

$$L(\boldsymbol{\phi}|\mathbf{D}_{\text{obs}}) = \prod_{i=1}^{n} f(\boldsymbol{y}^*_i, t_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{y}_i, \boldsymbol{z}_i). \tag{3.19}$$

4 Bayesian Inference

4.1 The priors and posterior

For $\boldsymbol{\theta}_0$, $\boldsymbol{\theta}_1$, $\boldsymbol{\gamma}$, α, and $\boldsymbol{\beta}$, we assume independent normal priors with $\boldsymbol{\theta}_k \sim \text{N}(\boldsymbol{\theta}_{\theta_k}, \boldsymbol{V}_{\theta_k})$ for $k = 0, 1$, $\boldsymbol{\gamma} \sim \text{N}(\boldsymbol{\theta}_\gamma, \boldsymbol{V}_\gamma)$, $\boldsymbol{\beta} \sim \text{N}(\boldsymbol{\theta}_\beta, \boldsymbol{V}_\beta)$, and $\alpha \sim \text{N}(\theta_\alpha, \sigma^2_\alpha)$ respectively, where $\boldsymbol{\theta}_{\theta_0} = \boldsymbol{\theta}_{\theta_1} = \boldsymbol{\theta}_\gamma = \theta_\alpha = \boldsymbol{\theta}_\beta = \mathbf{0}$, $\boldsymbol{V}_{\theta_0} = \boldsymbol{V}_{\theta_1} = 1000 I_2$, $\boldsymbol{V}_\gamma = \boldsymbol{V}_\beta = 1000 I_p$, and $\sigma^2_\alpha = 1000$. Independent gamma priors are assumed for $\boldsymbol{\lambda}$ with $\lambda_g \sim \text{Gamma}(a_g, b_g)$, where $a_g = 0.001$ and $b_g = 0.001$ for $g = 1, \ldots, G$. We assume an inverse gamma prior for the measurement errors with $\sigma^2 \sim \text{IG}(a_{\sigma^2}, b_{\sigma^2})$, where $a_{\sigma^2} = b_{\sigma^2} = 0.001$. Independent normal $N(0, 1)$ priors are assumed for $\log(\Gamma)$ on diagonal or Γ off diagonal with $\log(\Gamma_{11})$, Γ_{21}, $\log(\Gamma_{22})$, $\log(\Gamma_{33})$, Γ_{43}, and $\log(\Gamma_{44})$, respectively. The joint prior specified above is denoted by $\pi(\boldsymbol{\phi})$. The joint posterior distribution of $\boldsymbol{\phi}$ is

$$\pi(\boldsymbol{\phi}|\mathbf{D}_{\text{obs}}) = \frac{L(\boldsymbol{\phi}|\mathbf{D}_{\text{obs}})\pi(\boldsymbol{\phi})}{c(\text{D}_{\text{obs}})}, \tag{4.1}$$

where the normalizing constant is defined as

$$c(\text{D}_{\text{obs}}) = \int f(\boldsymbol{y}^*_i, t_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\pi(\boldsymbol{\phi}) d\boldsymbol{\phi}. \tag{4.2}$$

Let $\boldsymbol{\theta}^R = (\boldsymbol{\theta}^R_i, i = 1, \ldots, n)$. The joint posterior distribution of $(\boldsymbol{\phi}, \boldsymbol{\theta}^R)$ can be expressed as

$$\pi(\boldsymbol{\phi}, \boldsymbol{\theta}^R|\mathbf{D}_{\text{obs}}) = \frac{\prod_{i=1}^n f(\boldsymbol{y}^*_i, t_i, \boldsymbol{\theta}^R_i|\boldsymbol{\phi}, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{x}_i, \boldsymbol{z}_i)\pi(\boldsymbol{\phi})}{c(\text{D}_{\text{obs}})}. \tag{4.3}$$

4.2 Posterior sampling

Since the posterior distribution defined in (4.3) is analytically intractable, we develop a MCMC sampling algorithm to sample $(\boldsymbol{\phi}, \boldsymbol{\theta}^R, \boldsymbol{y}_i^*)$. To run the MCMC algorithm, we need to sample from the following conditional posterior distributions in turn:

(a) $[\boldsymbol{\theta}|\boldsymbol{\theta}_i^R, \boldsymbol{\gamma}, \boldsymbol{\sigma}^2, \Gamma, \boldsymbol{\lambda}, \alpha, \boldsymbol{\beta}, \mathrm{D}_{obs}]$ with localized Metropolis-Hasting algorithm;

(b) $[\boldsymbol{\gamma}|\boldsymbol{\theta}, \boldsymbol{\Gamma}, \boldsymbol{\theta}_i^R, \boldsymbol{\sigma}^2, \boldsymbol{\lambda}, \alpha, \boldsymbol{\beta}, \mathrm{D}_{obs}]$ with localized Metropolis-Hasting algorithm;

(c) $[\boldsymbol{\Gamma}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\theta}_i^R, \boldsymbol{\sigma}^2, \boldsymbol{\lambda}, \alpha, \boldsymbol{\beta}, \mathrm{D}_{obs}]$ with localized Metropolis-Hasting algorithm;

(d) $[\sigma^2|\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\theta}_i^R, \boldsymbol{\kappa}, \mathrm{D}_{obs}] \sim$ Inverse-Gamma distribution;

(e) $[\alpha|\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\theta}^R, \mathrm{D}_{obs}]$ with localized Metropolis-Hasting algorithm;

(f) $[\boldsymbol{\beta}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma^2, \Gamma, \boldsymbol{\theta}^R, \boldsymbol{\lambda}, \alpha, \mathrm{D}_{obs}]$ with localized Metropolis-Hasting algorithm;

(g) $[\lambda_g|\alpha, \boldsymbol{\beta}_k, \boldsymbol{\theta}, \boldsymbol{\gamma}, \Gamma, \boldsymbol{\theta}_i^R, \mathrm{D}_{obs}] \sim$ Gamma distribution;

(h) $[\boldsymbol{\theta}_i^R|\boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma^2, \Gamma, \boldsymbol{\lambda}, \alpha, \boldsymbol{\beta}, \mathrm{D}_{obs}]$ with localized Metropolis-Hasting algorithm;

(i) $[\boldsymbol{y}_i^*|\boldsymbol{y}_i, \boldsymbol{\theta}, \boldsymbol{\gamma}, \sigma^2, \Gamma, \boldsymbol{\theta}_i^R, \boldsymbol{\lambda}, \alpha, \boldsymbol{\beta}, \mathrm{D}_{obs}] \sim$ Truncated Normal distribution.

In order to evaluate the integral $H_i = \int_0^{t_i} h_0(u)\exp\{\alpha\mathcal{X}_i(u) + \boldsymbol{z}_i'\boldsymbol{\beta}\}du$ in (3.13), we use the Riemann integral approach. The details of the Riemann integral can be found in the Appendix.

4.3 Bayesian model assessment

Using the formula in (3.18), following Zhang et al. (2017), we can develop the DIC decomposition. However, the DIC decomposition requires the evaluation of four-dimensional integrals. Hence, we develop a variation of the DIC "decomposition", which does not require any numerical integration. Specifically, we define the deviance function as

$$\begin{aligned}\mathrm{Dev}(\boldsymbol{\phi}, \boldsymbol{\theta}^R) =& -2\sum_{i=1}^{n}\log(f(\boldsymbol{y}_i^*|\boldsymbol{\phi}_1, \boldsymbol{\theta}_i^R, \boldsymbol{\kappa}_i, \boldsymbol{x}_i)f(t_i|\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_i^R, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{z}_i)) \\ =& -2\sum_{i=1}^{n}\log(f(\boldsymbol{y}_i^*|\boldsymbol{\phi}_1, \boldsymbol{\theta}_i^R, \boldsymbol{\kappa}_i, \boldsymbol{x}_i)) - 2\sum_{i=1}^{n}\log(f(t_i|\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\theta}_i^R, \delta_i, \boldsymbol{\kappa}_i, \boldsymbol{z}_i)) \\ \equiv& \mathrm{Dev}_{\mathrm{Long}}(\boldsymbol{\phi}_1, \boldsymbol{\theta}^R) + \mathrm{Dev}_{\mathrm{Surv}}(\boldsymbol{\phi}_2, \boldsymbol{\theta}^R). \end{aligned} \tag{4.4}$$

Following the Deviance Information Criterion (DIC) defined in (Spiegelhalter et al. (2002)), we can construct $\mathrm{DIC}_{\mathrm{Long}}$ using $\mathrm{Dev}_{\mathrm{Long}}$ and $\mathrm{DIC}_{\mathrm{Surv}}$ using $\mathrm{Dev}_{\mathrm{Surv}}$. Mathematically, we have

$$\mathrm{DIC}_{\mathrm{Long}} = \mathrm{Dev}_{\mathrm{Long}}(\overline{\boldsymbol{\phi}_1}, \overline{\boldsymbol{\theta}^R}) + 2p_{D[\mathrm{Long}]}, \tag{4.5}$$

where $p_{D[\mathrm{Long}]} = E[\mathrm{Dev}_{\mathrm{Long}}(\boldsymbol{\phi}_1, \boldsymbol{\theta}^R)|\mathrm{D}_{\mathrm{obs}}] - \mathrm{Dev}_{\mathrm{Long}}(\overline{\boldsymbol{\phi}_1}, \overline{\boldsymbol{\theta}^R})$, and the expectation is taken with respect to the posterior distribution given in (4.3). The $\mathrm{DIC}_{\mathrm{Surv}}$ can be calculated in a similar fashion.

This variation of DIC can be computed using the MCMC samples from the posterior distribution under joint model. Therefore, this variation of DIC is very computationally efficient and convenient.

Under (4.4), the total DIC for the joint model has the following decomposition:

$$\text{DIC} = \text{DIC}_{\text{Long}} + \text{DIC}_{\text{Surv}}. \tag{4.6}$$

To quantify the improvement in the fit of the survival data by the joint model, we define $\Delta\text{DIC}_{\text{Surv}}$ as follows:

$$\Delta\text{DIC}_{\text{Surv}} = \text{DIC}_{\text{Surv},0} - \text{DIC}_{\text{Surv}}, \tag{4.7}$$

where $\text{DIC}_{\text{Surv},0}$ is defined under the survival-only model by letting the association parameters $\boldsymbol{\alpha}$ be a zero vector. Then, we compute the percentage of improvement from the survival-only model, which is denoted by $\%\Delta\text{DIC}_{\text{Surv}}$ and defined as

$$\%\Delta\text{DIC}_{\text{Surv}} = \frac{\Delta\text{DIC}_{\text{Surv}}}{\text{DIC}_{\text{Surv},0}}. \tag{4.8}$$

4.4 Concordance for ordinal responses

We develop a variation of concordance to evaluate the longitudinal model performance for ordinal data. Concordance evaluates the predictive accuracy of a model by measuring the agreement between observed and predicted outcomes, which is the number of concordant pair divided by the total number of comparable pairs.

In the ordinal response setting, using the notation in Section 3.1, for the ith subject, a within-subject pair, (i, j) and (i, j'), where $j < j'$, is comparable if $y_i(a_{ij}) \neq y_i(a_{ij'})$; and for subjects i and i', where $i < i'$, a between-subject pair, (i, j) and (i', j'), is comparable if $y_i(a_{ij}) \neq y_{i'}(a_{i'j'})$, where $1 \leq j \leq m_i$ and $1 \leq j' \leq m_{i'}$. Furthermore, a comparable within-subject pair, (i, j) and (i, j'), is concordant if

$$\begin{aligned}
&y_i(a_{ij}) > y_i(a_{ij'}) \text{ and} \\
&\theta_{\kappa_i(a_{ij})0} + \theta_{\kappa_i(a_{ij})0i} + (\theta_{\kappa_i(a_{ij})1} + \theta_{\kappa_i(a_{ij})1i})t^*(a_{ij}) > \\
&\theta_{\kappa_i(a_{ij'})0} + \theta_{\kappa_i(a_{ij'})0i} + (\theta_{\kappa_i(a_{ij'})1} + \theta_{\kappa_i(a_{ij'})1i})t^*(a_{ij'})
\end{aligned} \tag{4.9}$$

or

$$\begin{aligned}
&y_i(a_{ij}) < y_i(a_{ij'}) \text{ and} \\
&\theta_{\kappa_i(a_{ij})0} + \theta_{\kappa_i(a_{ij})0i} + (\theta_{\kappa_i(a_{ij})1} + \theta_{\kappa_i(a_{ij})1i})t^*(a_{ij}) < \\
&\theta_{\kappa_i(a_{ij'})0} + \theta_{\kappa_i(a_{ij'})0i} + (\theta_{\kappa_i(a_{ij'})1} + \theta_{\kappa_i(a_{ij'})1i})t^*(a_{ij'}).
\end{aligned} \tag{4.10}$$

Similarly, a comparable between-subject pair, (i, j) and (i', j'), is concordant if

$$\begin{aligned}
&y_i(a_{ij}) > y_i(a_{i'j'}) \text{ and} \\
&\theta_{\kappa_i(a_{ij})0} + \theta_{\kappa_i(a_{ij})0i} + (\theta_{\kappa_i(a_{ij})1} + \theta_{\kappa_i(a_{ij})1i})t^*(a_{ij}) + \boldsymbol{x}_i'\boldsymbol{\gamma} > \\
&\theta_{\kappa_i(a_{i'j'})0} + \theta_{\kappa_i(a_{i'j'})0i} + (\theta_{\kappa_i(a_{i'j'})1} + \theta_{\kappa_i(a_{i'j'})1i})t^*(a_{i'j'}) + \boldsymbol{x}_{i'}'\boldsymbol{\gamma}
\end{aligned} \tag{4.11}$$

or

$$\begin{aligned}
&y_i(a_{ij}) < y_i(a_{i'j'}) \text{ and} \\
&\theta_{\kappa_i(a_{ij})0} + \theta_{\kappa_i(a_{ij})0i} + (\theta_{\kappa_i(a_{ij})1} + \theta_{\kappa_i(a_{ij})1i})t^*(a_{ij}) + \boldsymbol{x}_i'\boldsymbol{\gamma} < \\
&\theta_{\kappa_i(a_{i'j'})0} + \theta_{\kappa_i(a_{i'j'})0i} + (\theta_{\kappa_i(a_{i'j'})1} + \theta_{\kappa_i(a_{i'j'})1i})t^*(a_{i'j'}) + \boldsymbol{x}_{i'}'\boldsymbol{\gamma}.
\end{aligned} \tag{4.12}$$

A comparable pair is called a tie if the second inequality ($<$ or $>$) in (4.9) – (4.12) becomes an equality, "=". Then we can define the concordance for the ordinal data as

$$C = \frac{\text{total number of condance pairs} + 0.5 \times \text{total number of ties}}{\text{total number of comparable pairs}}. \tag{4.13}$$

Subsequently, we compute the percentage of improvement from the conventional linear mixed-effects model, which is denoted by $\%\Delta\text{C}_{\text{Long}}$ and defined as

$$\%\Delta\text{C}_{\text{Long}} = \frac{\Delta\text{C}_{\text{Long}}}{\text{C}_{\text{LMM}}}. \tag{4.14}$$

5 Analysis of the ADNI Data

We analyze the ADNI data discussed in Section 2 using the proposed methods for ordinal data. The longitudinal measurement we analyze is the Clinical Dementia Rating (CDR). We executed 5000 iterations for burn-in, and subsequently collected 10,000 MCMC samples. The posterior estimates and 95% highest posterior density (HPD) intervals, as outlined by Chen and Shao (1999), are detailed in Tables 3 and 4. For the survival sub-model, we select the optimal number of time intervals based on the DIC variations. The survival sub-model with 9 pieces yields the smallest $\text{DIC}_{\text{Surv|Long}}$. For computational stability, all continuous covariates are standardized in the analysis.

The intercept of the trajectory when the patient is in CN is statistically significantly lower than the intercept when the patient is in MCI. Covariates including marital status, APOE4 and RAVLT show a positive association with CDR values. Individuals who are married with a higher APOE4 count and higher RAVLT forgetting percentages tend to exhibit a significantly higher CDR. The association parameters are also significantly different between states, indicating a different effect of CDR on progression to AD when the patient is in different period of disease progression.

Table 5 displays the DIC decomposition results, providing an assessment of how the inclusion of longitudinal data in the survival sub-model influences the fit of survival outcomes. The $\%\Delta\text{DIC}_{\text{Surv}}$ draw the conclusion that the Clinical Dementia Rating (CDR) contributes significantly to the improvement of survival fit by 7.37%.

To assess the performance of the proposed state-specific trajectory model, we evaluate the DIC variations and the concordance of SSTM and Linear Mixed Model with interaction of state and time. We quantify the marginal improvement of the longitudinal model by implementing SSTM over LMM and also the conditional improvement of the fit of the survival data by including the longitudinal biomarker using SSTM over LMM. According to Table 6, the fit of the marginal longitudinal model is improved by 5.82% and 13.50% in terms of DIC and concordance, respectively. The fit of the survival data given longitudinal biomarker is improved by 4.38%, according to DIC.

Table 3: Posterior estimates of the longitudinal parameters for CDR (higher means worse) under SSTM

Para.	Post. mean	Post. SD	95% HPD interval
$\theta_{\text{CN},0}$	1.0356	0.0246	(0.9842, 1.0829)
$\theta_{\text{CN},1}$	0.0376	0.0042	(0.0289, 0.0458)
$\theta_{\text{MCI},0}$	1.9632	0.0229	(1.9175, 2.0066)
$\theta_{\text{MCI},1}$	0.0409	0.0053	(0.0307, 0.0511)
$\theta_{\text{CN},0} - \theta_{\text{MCI},0}$	-0.9276	0.0269	(-0.9854, -0.8769)
$\theta_{\text{CN},1} - \theta_{\text{MCI},1}$	-0.0033	0.0068	(-0.0163, 0.0107)
γ_{age}	0.0053	0.0036	(-0.0021, 0.0119)
γ_{sex}	-0.0025	0.0075	(-0.0171, 0.0124)
γ_{race}	0.0189	0.0141	(-0.0074, 0.0471)
γ_{marital}	0.0210	0.0083	(0.0052, 0.0374)
$\gamma_{\text{education}}$	-0.0029	0.0035	(-0.0100, 0.0035)
γ_{APOE4}	0.0196	0.0058	(0.0084, 0.0311)
γ_{RAVLT}	0.0152	0.0038	(0.0083, 0.0230)
b_{11}	0.0194	0.0064	(0.0072, 0.0349)
b_{21}	0.0005	0.0020	(-0.0036, 0.0041)
b_{22}	0.0045	0.0011	(0.0019, 0.0064)
b_{33}	0.0263	0.0035	(0.0200, 0.0334)
b_{43}	0.0035	0.0013	(0.0007, 0.0059)
b_{44}	0.0043	0.0018	(0.0001, 0.0071)
$\Omega_{11} - \Omega_{33}$	-0.0003	0.0003	(-0.0009, 0.0004)
$\Omega_{21} - \Omega_{43}$	-0.0001	0.0000	(-0.0002, 0.0000)
$\Omega_{22} - \Omega_{44}$	0.0000	0.0000	(0.0000, 0.0000)
σ^2	0.0800	0.0013	(0.0776, 0.0827)

Table 4: Posterior estimates of the survival parameters for the association between CDR and time to AD

	Posterior mean	Posterior SD	95% HPD interval
α_{CN}	2.0171	0.1444	(1.7313, 2.2895)
α_{MCI}	-0.0649	0.0731	(-0.1965, 0.0796)
$\alpha_{\text{CN}} - \alpha_{\text{MCI}}$	2.0820	0.0856	(1.9226, 2.2590)
β_{age}	0.2990	0.0533	(0.1983, 0.4067)
β_{sex}	-0.0207	0.1136	(-0.2465, 0.1969)
β_{race}	-0.4635	0.1416	(-0.7219, -0.1777)
β_{marital}	-0.0907	0.1198	(-0.3178, 0.1561)
$\beta_{\text{education}}$	-0.0137	0.0505	(-0.1198, 0.0796)
β_{APOE4}	0.4089	0.0754	(0.2694, 0.5634)
β_{RAVLT}	0.4313	0.0571	(0.3277, 0.5523)

Table 5: $\Delta\text{DIC}_{\text{Surv}}$

Response	$\text{DIC}_{\text{Surv},0}$	$\text{DIC}_{\text{Surv}\|\text{Long}}$	$\Delta\text{DIC}_{\text{Surv}}$	$\%\Delta\text{DIC}_{\text{Surv}}$
CDR	602.7863	558.3423	44.4440	7.37%

Table 6: $\Delta\text{DIC}_{\text{Long}}$, $\Delta\text{DIC}_{\text{Surv}}$, and $\Delta\text{C}_{\text{Long}}$ of CDR

Response	CDR	Response	CDR	Response	CDR
DIC_{LMM}	11857.3477	$\text{DIC}_{\text{Surv}\|\text{LMM}}$	583.9144	C_{LMM}	0.8022
DIC_{SSTM}	11167.3224	$\text{DIC}_{\text{Surv}\|\text{SSTM}}$	558.3423	C_{SSTM}	0.9105
$\Delta\text{DIC}_{\text{Long}}$	690.0253	$\Delta\text{DIC}_{\text{Surv}}$	25.5721	$\Delta\text{C}_{\text{Long}}$	0.1083
$\%\Delta\text{DIC}_{\text{Long}}$	5.82%	$\%\Delta\text{DIC}_{\text{Surv}}$	4.38%	$\%\Delta\text{C}_{\text{Long}}$	13.50%

6 Discussion

In this study, we introduce the joint model for longitudinal ordinal data with state-specific trajectory and time-to-event data, demonstrating its applicability to Alzheimer's Disease. This model captures the fluctuation of cognitive condition before and after state transition, enhancing the longitudinal fit which is quantified by the proposed C-index. Additionally, the joint model incorporates the contribution of longitudinal data into the survival sub-model. To quantify the improved fit of the survival sub-model from the inclusion of longitudinal data, we calculated $\Delta\text{DIC}_{\text{Surv}}$. These results indicate a substantial enhancement in terms of survival fit. The computation for this paper is carried out using FORTRAN compiler and the IMSL FORTRAN library, with double precision accuracy. The FORTRAN code can be provided by the authors upon request.

Following are some potential future research developments based on this study: (i) we could consider relaxing the assumption of independence between the random effects among states; (ii) the transition time could be modeled using hidden Markov model; (iii) develop the decomposition of WAIC by integrating out the random effects from the joint posterior instead of treating the random effects as parameters; and (iv) jointly model the longitudinal biomarker and multi-state transition, by employing a multi-state Markov transition model to study the transitions across all three states, providing a comprehensive view of the disease progression and its correlation with various risk factors.

Acknowledgements

The authors thank Dr. Abidemi Adeniji for introducing the motivating data and helpful suggestions. The authors would also like to thank the Editors and the reviewer for their helpful comments, which have led to an improved version of the manuscript.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number
W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the Uni-

versity of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

A Appendix

A.1 Proposed state-specific trajectory Model when the biomarker and diagnosis are examined in separate visits

Suppose we have m_i longitudinal measurements and r_i diagnosis results observed for the i-th subject for $i = 1, \ldots, n$. Let a_{ij} denote the time in years of the j-th biomarker visit from the baseline for $j = 0, 1, \ldots, m_i$ and $i = 1, \ldots, n$. Let $a^*_{i\ell}$ denote the time in years of the ℓ-th diagnosis visit from the baseline for $\ell = 0, 1, \ldots, r_i$ and $i = 1, \ldots, n$. Denote $\mathbf{y}_i(a_{ij})$ as the longitudinal measurement for subject i at time a_{ij}. The diagnosed states of subject i at time a_{ij} and $a^*_{i\ell}$ are represented by $\kappa_i(a_{ij})$ and $\kappa_i(a^*_{i\ell})$, respectively. Specifically, $\kappa_i(a_{ij})$ equals 0 if subject i is diagnosed as Cognitive Normal (CN) at time a_{ij} and 1 if diagnosed as Mild Cognitive Impairment (MCI). Assume $a_{i,-1} = 0$. The longitudinal model is defined as

$$\begin{aligned}
\mathbf{y}_i(a_{ij}) = \sum_{\ell: a_{i,j-1} < a^*_{i\ell} \le a_{ij}} & \Big\{ (\theta_{\kappa_i(a^*_{i\ell})0} + \theta_{\kappa_i(a^*_{i\ell})0i}) \\
& \times 1\{(\ell = \min\{\ell : a_{i,j-1} < a^*_{i\ell} \le a_{ij}\}) \cup (\kappa_i(a^*_{i\ell}) \neq \kappa_i(a_{i,\ell-1}))\} \\
& + (\theta_{\kappa_i(a^*_{i\ell})1} + \theta_{\kappa_i(a^*_{i\ell})1i}) t^*_{\kappa_i(a^*_{i\ell})}(a^*_{i\ell}) \Big\} + (\theta_{\kappa_i(a_{ij})0} + \theta_{\kappa_i(a_{ij})0i}) \\
& \times 1\{(\{\ell : a_{i,j-1} < a^*_{i\ell} \le a_{ij}\} = \emptyset) \cup (\kappa_i(a_{ij}) \neq \kappa_i(\max\{a^*_{i\ell} : a^*_{i\ell} \le a_{ij}\}))\} \\
& + (\theta_{\kappa_i(\max\{a^*_{i\ell}: a^*_{i\ell} \le a_{ij}\})1} + \theta_{\kappa_i(\max\{a^*_{i\ell}: a^*_{i\ell} \le a_{ij}\})1i}) t^*_{\kappa_i(\max\{a^*_{i\ell}: a^*_{i\ell} \le a_{ij}\})}(a_{ij}) + \mathbf{x}_i'\boldsymbol{\gamma} + \boldsymbol{\epsilon}_{ij},
\end{aligned} \tag{A.1}$$

where $\theta_{\kappa 0}$ and $\theta_{\kappa 1}$ are the fixed intercept and fixed slope, respectively, for the subject who is diagnosed as state κ, $\theta_{\kappa 0i}$ and $\theta_{\kappa 1i}$ are the state-specific and subject-specific random intercept and random slope, respectively, $\boldsymbol{\gamma}$ is a p-dimensional vector of coefficients corresponding to the p-dimensional covariates $\boldsymbol{x}_i$, $t^*(a^*_{i\ell})$ and $t^*(a_{ij})$ are the cumulative time that subject i stays in the current state until time $a^*_{i\ell}$ and a_{ij}, respectively, and they are defined as

$$t^*(a^*_{i\ell}) = \begin{cases} 0, & \text{for } \ell = 0, \\ a^*_{i\ell} - \max\{a_{i\ell^*} 1\{\kappa_i(a_{i\ell^*}) = 1 - \kappa_i(a_{ij})\}, \ell^* = 0, \ldots, \ell - 1\}, & \text{for } \ell = 1, \ldots, r_i, \text{and} \end{cases} \tag{A.2}$$

$$t^*(a_{ij}) = \begin{cases} 0, & \text{for } j = 0, \\ a_{ij} - \max\{a^*_{i\ell} 1\{\kappa_i(a^*_{i\ell}) = 1 - \kappa_i(a_{ij})\} : a^*_{i\ell} \le a_{ij}\}, & \text{for } j = 1, 2, \ldots, m_i. \end{cases} \tag{A.3}$$

The error term $\boldsymbol{\epsilon}_{ij}$ is assumed to follow $N(0, \sigma^2)$. The random effects and error term are assumed to be independent.

A.2 Riemann integral to evaluate the integration in the survival sub-model

In order to evaluate the integral $H_i = \int_0^{t_i} h_0(u)\exp\{\alpha\mathcal{X}_i(u) + \boldsymbol{z}_i'\boldsymbol{\beta}\}du$ in (3.13), we use the Riemann integral approach. Let $a^*_{i\ell} = \alpha(\theta_{i,2\kappa^*_{i\ell}+1} + \boldsymbol{x}_i'\boldsymbol{\gamma})$, $B_{i\ell} = \alpha\theta_{i,2\kappa^*_{i\ell}+2}$, and $C_{i\ell} = a^*_{i\ell} - B_{i\ell}\theta^*_{i\ell} + \boldsymbol{z}_i'\boldsymbol{\beta}$.

Let NS_i be the number of times the ith subject switches the state. For example, if a subject stay in CN the whole time, then $\text{NS}_i = 0$. If a subject starts with CN, switches to MCI and later switches back to CN, then $\text{NS}_i = 0$. Let θ_{ih} denote a vector of time points consisting of time at baseline, 0, the time point when the i-th subject switches state and the event or censored time. $\mathcal{I}_{i\ell}$ represents the ℓ-th time interval, $\ell = 1, \ldots, \text{NS}_i + 1$, segmented by the state switch. And $\kappa^*_{i\ell}$ equals 0 if the i-th subject is in CN during the time interval $\mathcal{I}_{i\ell}$; 1 if in MCI. Let $g_l = \text{argmax}_g\{s_g | s_g \in \mathcal{I}_{i\ell}\}$ denote the last time point of $\boldsymbol{s}$ in the interval $\mathcal{I}_{i\ell}$ with $g_0 = 0$. If there does not exist any s_g within an interval, then $g_l = g_{l-1}$.

For example, if a patient is in CN in year [0,2), MCI in [2,4), CN in [4,7) and MCI in [7,8], where 8 is the censored time. Then $\text{NS}_i = 3$, $\theta_i = (0, 2, 4, 7, 8)$, and $\mathcal{I}_{i1} = [0, 2)$, $\kappa_{i1} = 0$, $\mathcal{I}_{i2} = [2, 4)$, $\kappa_{i2} = 1$, $\mathcal{I}_{i3} = [4, 7)$, $\kappa_{i3} = 0$, $\mathcal{I}_{i4} = [7, 8]$, $\kappa_{i4} = 1$.

Let N_1, N_2, N_3 denote the numbers of segments in the corresponding three parts of the Riemann integral. We have

$$
\begin{aligned}
&\int_0^{t_i} h_0(u)\exp\{\alpha\mathcal{X}_i(u) + \boldsymbol{z}_i'\boldsymbol{\beta}\}du \qquad (A.4)\\
&= \sum_{l=1}^{\text{NS}_i+1} \Bigg\{ 1\{g_{l-1} \neq g_l\}\Bigg\{ \int_{\theta^*_{i\ell}}^{s_{g_{l-1}+1}} \lambda_{g_{l-1}+1}\exp\{B_{i\ell}u + C_{i\ell}\}du + 1\{g_{l-1}+1 \neq g_l\}\\
&\quad \times \Bigg[\sum_{g=g_{l-1}+1}^{g_l} \int_{s_{g-1}}^{s_g} \lambda_g\exp\{B_{i\ell}u + C_{i\ell}\}du \Bigg] \Bigg\} + \int_{\max\{s_{g_l}, \theta^*_{i\ell}\}}^{\theta_{i,l+1}} \lambda_{g_l+1}\exp\{B_{i\ell}u + C_{i\ell}\}du \Bigg\}\\
&= \sum_{l=1}^{\text{NS}_i+1} \Bigg\{ 1\{g_{l-1} \neq g_l\}\Bigg\{ \sum_{k=1}^{N_1} \lambda_{g_{l-1}+1}\Big(\frac{s_{g_{l-1}+1} - \theta^*_{i\ell}}{N_1}\Big)\exp\Bigg\{ \frac{(2k-1)B_{i\ell}(s_{g_{l-1}+1} - \theta^*_{i\ell})}{2N_1} + C_{i\ell} \Bigg\}\\
&\quad + 1\{g_{l-1}+1 \neq g_l\}\Bigg[\sum_{g=g_{l-1}+1}^{g_l} \sum_{k=1}^{N_2} \lambda_g\Big(\frac{s_g - s_{g-1}}{N_2}\Big)\exp\Bigg\{ \frac{(2k-1)B_{i\ell}(s_g - s_{g-1})}{2N_2} + C_{i\ell} \Bigg\} \Bigg] \Bigg\}\\
&\quad + \sum_{k=1}^{N_3} \lambda_{g_l+1}\Big(\frac{\theta_{i,l+1} - \max\{s_{g_l}, \theta^*_{i\ell}\}}{N_3}\Big)\exp\Bigg\{ \frac{(2k-1)B_{i\ell}(\theta_{i,l+1} - \max\{s_{g_l}, \theta^*_{i\ell}\})}{2N_3} + C_{i\ell} \Bigg\} \Bigg\}.
\end{aligned}
$$

Algorithm to compute the integral in (A.4):

Step 1. Divide the trajectory into $\text{NS}_i + 1$ intervals separated by state transitions.

Step 2. In each interval, determine $\text{NG}_{i\ell}$, the number of time points s_g within the interval $\mathcal{I}_{i\ell}$.

Step 3. If $\text{NG}_{i\ell} = 0$, then go to Step 4; else if $\text{NG}_{i\ell} = 1$, then go to Step 5; otherwise, go to Step 6.

Step 4. The integral over $\mathcal{I}_{i\ell}$ can be evaluated by

$$\lambda_{g_l+1}\Big(\frac{\theta_{i,l+1}-\theta^*_{i\ell}}{N_3}\Big)\sum_{k=1}^{N_3}\exp\Big\{\frac{(2k-1)B_{i\ell}(\theta_{i,l+1}-\theta^*_{i\ell})}{2N_3}+C_{i\ell}\Big\}.$$

Step 5. The integral over $\mathcal{I}_{i\ell}$ can be evaluated by

$$\lambda_{g_l}\Big(\frac{s_{g_l}-\theta^*_{i\ell}}{N_1}\Big)\sum_{k=1}^{N_1}\exp\Big\{\frac{(2k-1)B_{i\ell}(s_{g_l}-\theta^*_{i\ell})}{2N_1}+C_{i\ell}\Big\}+$$
$$\lambda_{g_l+1}\Big(\frac{\theta_{i,l+1}-s_{g_l}}{N_3}\Big)\sum_{k=1}^{N_3}\exp\Big\{\frac{(2k-1)B_{i\ell}(\theta_{i,l+1}-s_{g_l})}{2N_3}+C_{i\ell}\Big\}.$$

Step 6. The integral over $\mathcal{I}_{i\ell}$ can be evaluated by

$$\lambda_{g_{l-1}+1}\Big(\frac{s_{g_{l-1}+1}-\theta^*_{i\ell}}{N_1}\Big)\sum_{k=1}^{N_1}\exp\Big\{\frac{(2k-1)B_{i\ell}(s_{g_{l-1}+1}-\theta^*_{i\ell})}{2N_1}+C_{i\ell}\Big\}+$$
$$\sum_{g=g_{l-1}+1}^{g_l}\lambda_g\Big(\frac{s_g-s_{g-1}}{N_2}\Big)\sum_{k=1}^{N_2}\exp\Big\{\frac{(2k-1)B_{i\ell}(s_g-s_{g-1})}{2N_2}+C_{i\ell}\Big\}+$$
$$\lambda_{g_l+1}\Big(\frac{\theta_{i,l+1}-s_{g_l}}{N_3}\Big)\sum_{k=1}^{N_3}\exp\Big\{\frac{(2k-1)B_{i\ell}(\theta_{i,l+1}-s_{g_l})}{2N_3}+C_{i\ell}\Big\}.$$

Step 7. Evaluate the integral in (A.4) by summing up the integration results on all intervals.

References

Albert, J. H. and Chib, S. (1993), "Bayesian analysis of binary and polychotomous response data," *Journal of the American statistical Association*, 88, 669–679.

Andrinopoulou, E.-R., Harhay, M. O., Ratcliffe, S. J., and Rizopoulos, D. (2021), "Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data," *International Journal of Epidemiology*, 50, 1731–1743.

Bernal-Rusiel, J. L., Greve, D. N., Reuter, M., Fischl, B., Sabuncu, M. R., Initiative, A. D. N., et al. (2013a), "Statistical analysis of longitudinal neuroimage data with linear mixed effects models," *Neuroimage*, 66, 249–260.

Bernal-Rusiel, J. L., Reuter, M., Greve, D. N., Fischl, B., Sabuncu, M. R., Initiative, A. D. N., et al. (2013b), "Spatiotemporal linear mixed effects modeling for the mass-univariate analysis of longitudinal neuroimage data," *Neuroimage*, 81, 358–370.

Chen, M.-H., Ibrahim, J. G., and Sinha, D. (2004), "A new joint model for longitudinal and survival data with a cure fraction," *Journal of Multivariate Analysis*, 91, 18–34.

Chen, M.-H. and Shao, Q.-M. (1999), "Monte Carlo estimation of Bayesian credible and HPD intervals," *Journal of computational and Graphical Statistics*, 8, 69–92.

Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000), *Monte Carlo methods in Bayesian computation*, Springer Science & Business Media.

Cowles, M. K., Carlin, B. P., and Connett, J. E. (1996), "Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness," *Journal of the American Statistical Association*, 91, 86–98.

Cox, D. R. (1972), "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 34, 187–202.

Cummings, J. (2023), "Meaningful benefit and minimal clinically important difference (MCID) in Alzheimer's disease: Open peer commentary," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 9, e12411.

Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D., (ADNI), A. D. N. I., et al. (2016), "Instantiated mixed effects modeling of Alzheimer's disease markers," *NeuroImage*, 142, 113–125.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*.

Hedeker, D. and Gibbons, R. D. (1994), "A random-effects ordinal regression model for multilevel analysis," *Biometrics*, 933–944.

Hu, W., Li, G., and Li, N. (2009), "A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data," *Statistics in medicine*, 28, 1601–1619.

Huang, X., Li, G., Elashoff, R. M., and Pan, J. (2011), "A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects," *Lifetime data analysis*, 17, 80.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2004), "Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials," *Statistica Sinica*, 863–883.

Jacqmin-Gadda, H., Proust-Lima, C., and Amiéva, H. (2010), "Semi-parametric latent process model for longitudinal ordinal data: Application to cognitive decline," *Statistics in Medicine*, 29, 2723–2731.

Laird, N. M. and Ware, J. H. (1982), "Random-effects models for longitudinal data," *Biometrics*, 963–974.

Lee, K. and Daniels, M. J. (2008), "Marginalized models for longitudinal ordinal data with application to quality of life studies," *Statistics in Medicine*, 27, 4359–4380.

Li, C., Xiao, L., and Luo, S. (2022), "Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's Disease," *Biometrics*, 78, 435–447.

Li, D., Iddi, S., Thompson, W. K., Donohue, M. C., and Initiative, A. D. N. (2019), "Bayesian latent time joint mixed effect models for multicohort longitudinal data," *Statistical methods in medical research*, 28, 835–845.

Li, D., Iddi, S., Thompson, W. K., Rafii, M. S., Aisen, P. S., Donohue, M. C., and Initiative, A. D. N. (2018a), "Bayesian latent time joint mixed-effects model of progression in the Alzheimer's Disease Neuroimaging Initiative," *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 657–668.

Li, K., Chan, W., Doody, R. S., Quinn, J., Luo, S., Initiative, A. D. N., et al. (2017), "Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data," *Journal of Alzheimer's Disease*, 58, 361–371.

Li, K., O'Brien, R., Lutz, M., Luo, S., Initiative, A. D. N., et al. (2018b), "A prognostic model of Alzheimer's disease relying on multiple longitudinal measures and time-to-event data," *Alzheimer's & Dementia*, 14, 644–651.

Li, N., Elashoff, R. M., Li, G., and Saver, J. (2010), "Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial," *Statistics in medicine*, 29, 546–557.

Li, N., Liu, Y., Li, S., Elashoff, R. M., and Li, G. (2021), "A flexible joint model for multiple longitudinal biomarkers and a time-to-event outcome: With applications to dynamic prediction using highly correlated biomarkers," *Biometrical Journal*, 63, 1575–1586.

Lin, D. Y. and Wei, L.-J. (1989), "The robust inference for the Cox proportional hazards model," *Journal of the American statistical Association*, 84, 1074–1078.

Liu, L. C. and Hedeker, D. (2006), "A mixed-effects regression model for longitudinal multivariate ordinal data," *Biometrics*, 62, 261–268.

McCullagh, P. (1980), "Regression models for ordinal data," *Journal of the Royal Statistical Society: Series B (Methodological)*, 42, 109–127.

Medina-Olivares, V., Lindgren, F., Calabrese, R., and Crook, J. (2023), "Joint models of multivariate longitudinal outcomes and discrete survival data with INLA: An application to credit repayment behaviour," *European Journal of Operational Research*.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equation of state calculations by fast computing machines," *The journal of chemical physics*, 21, 1087–1092.

Molenberghs, G., Kenward, M. G., and Lesaffre, E. (1997), "The analysis of longitudinal ordinal data with nonrandom drop-out," *Biometrika*, 84, 33–44.

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, 223–240.

Sheikh, M. T., Chen, M.-H., Gelfond, J. A., and Ibrahim, J. G. (2022), "A Power Prior Approach for Leveraging External Longitudinal and Competing Risks Survival Data Within the Joint Modeling Framework," *Statistics in Biosciences*, 14, 318–336.

Sheikh, M. T., Ibrahim, J. G., Gelfond, J. A., Sun, W., and Chen, M.-H. (2021), "Joint modelling of longitudinal and survival data in the presence of competing risks with applications to prostate cancer data," *Statistical modelling*, 21, 72–94.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002), "Bayesian measures of model complexity and fit," *Journal of the royal statistical society: Series b (statistical methodology)*, 64, 583–639.

Tsiatis, A. A. and Davidian, M. (2004), "Joint modeling of longitudinal and time-to-event data: an overview," *Statistica Sinica*, 809–834.

Varin, C. and Czado, C. (2010), "A mixed autoregressive probit model for ordinal longitudinal data," *Biostatistics*, 11, 127–138.

Williams, M. M., Storandt, M., Roe, C. M., and Morris, J. C. (2013), "Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores," *Alzheimer's & Dementia*, 9, S39–S44.

Wu, Y., Zhang, X., He, Y., Cui, J., Ge, X., Han, H., Luo, Y., Liu, L., Wang, X., Yu, H., et al. (2020), "Predicting Alzheimer's disease based on survival data and longitudinally measured performance on cognitive and functional scales," *Psychiatry research*, 291, 113201.

Yi, J. and Tang, N. (2022), "Variational Bayesian inference in high-dimensional linear mixed models," *Mathematics*, 10, 463.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., and Shen, W. (2016), "JMFit: a SAS macro for joint models of longitudinal and survival data," *Journal of statistical software*, 71.

— (2017), "Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials," *Journal of Computational and Graphical Statistics*, 26, 121–133.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., Wang, P., and Shen, W. (2014), "Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials," *Statistics in Medicine*, 33, 4715–4733.

Zhang, F., Chen, M.-H., Cong, X. J., and Chen, Q. (2021), "Assessing importance of biomarkers: A Bayesian joint modelling approach of longitudinal and survival data with semi-competing risks," *Statistical modelling*, 21, 30–55.

Ziegler, G., Penny, W. D., Ridgway, G. R., Ourselin, S., Friston, K. J., Initiative, A. D. N., et al. (2015), "Estimating anatomical trajectories with Bayesian mixed-effects modeling," *Neuroimage*, 121, 51–68.

Received: February 17, 2024

Accepted: June 18, 2024