Probabilistic Clustering using Shared Latent Variable Model for Assessing Alzheimer's Disease Biomarkers

Yizhen Xu*, Zheyu Wang

1 Abstract

The pathophysiological process of Alzheimer's disease (AD) is not easily observable and may initiate a decade or more before symptoms appear. This work aims to quantify the AD pathophysiological pathway by a single latent trait metric, discover the sequential ordering of biomarkers along the metric, and identify subgroups with systematically different disease risk profiles. These goals are important for developing therapy and early AD detection strategy during the presymptomatic stages.

We propose a latent variable mixture model in which the components are structural equation models with a subject-specific continuous latent disease process (LDP), and subject-level clusters are identified based on systematically different profiles of unobserved heterogeneous disease progression, i.e. shift in the LDP. Model estimation uses Hamiltonian Monte Carlo and is validated by simulation studies. The number of clusters is determined by BIC. When applied to the BIOCARD (Biomarkers of Cognitive Decline Among Normal Individuals) data, the estimated sequential order of biomarkers is consistent with the hypothetical model of biomarker dynamics in Jack Jr et al. [2013]. Application results are further evaluated by investigating the conversion between AD clinical diagnoses within posterior clusters.

2 Introduction

Alzheimer's disease (AD) disrupts neuron communication in areas of the brain that control memory, language, and thought. It is a slow-progressing neurodegenerative disease that leads to a gradual deterioration of brain function. The time elapsed between the onset of AD-related pathophysiological changes and the occurrence of AD clinical symptoms is believed to be a decade or more, during which cognitive impairment accumulates until there is enough damage to warrant a clinical diagnosis [Jack Jr et al., 2010, 2013]. Furthermore, by 2050, the 65-and-older American population is expected to reach 88 million, with AD affecting 12.7 million people, imposing a total annual care-related payment of just under $1 trillion[Association et al., 2022]. Because AD is irreversible and has significant health costs, it is critical for AD research to focus on designing clinical trials for early interventions and developing procedures for early AD detection in the preclinical stage, before any symptoms appear [Dubois et al., 2016, Vermunt et al., 2019]. The difficulties stem from the inability to reliably diagnose AD in clinical settings, as well as a lack of an established connection between the dynamics of any specific biomarker during the asymptomatic phase and the subsequent manifestation of AD symptoms.

For AD to be ascertained, a neuropathological brain autopsy is needed, which requires adequate medical resources and can only be done after the person has died and given permission, while patient death usually occurs at least 5 years after a clinical diagnosis of AD [Jack Jr et al., 2010, 2013]. As a result of the absence of appropriate diagnostic tools and ethical concerns, the true degree of cognitive impairment is typically unobservable, and it is critical to employ repeated biomarker assessments from closely followed individuals to uncover a potential gold standard for diagnosing AD. Currently, the progression of AD is clinically characterized by three successive stages based on measurable biomarkers

and clinician judgement: cognitively normal (CN), mild cognitive impairment (MCI), and AD [Bennett et al., 2002, Petersen et al., 2009]. Sperling et al. [2011] suggested that asymptomatic subjects with AD-pathological processes may show subtle evidence of impairment in their biomarker measurements, and that both the pathological processes of AD and its clinical symptoms are best thought of as continuous processes.

The primary objective of this study is to investigate the temporal order of biomarkers with regard to a continous latent disease process (LDP). The discovery of the temporal ordering of biomarkers is important for two reasons: first, it facilitates finer disease staging, allowing therapies to be evaluated at earlier, pre-symptomatic stages, when interventions are most likely to be effective [Hampel et al., 2008], and second, it is required for understanding the dynamics of disease progression for early detection of the disease. Multiple studies developed conceptual frameworks and computational techniques to assess the dynamics of AD-related biomarkers. Jack Jr et al. [2010] hypothesized that the abnormality of biomarkers increases in a temporally ordered manner in relation to the discrete AD clinical diagnosis stages, with the rate of change being a sigmoid function. To test the hypotheses derived from Jack Jr et al. [2010], Han et al. [2012] fitted a sequence of linear mixed effects models on different classes of biomarkers. Following the conceptual construct in Jack Jr et al. [2010], the unsupervised machine learning method in Wang et al. [2020] and Wang and Zhou [2018] quantifies unobserved disease severity using ordinal class membership in the finite mixture model of continuous biomarkers.

Jack Jr et al. [2013] updated the conceptual framework in Jack Jr et al. [2010] so that, rather than discrete clinical diagnosis stages, biomarker changes are ordered sequentially along a continuum of AD represented by time. Based on this construct of AD pathology, Donohue et al. [2014], Li et al. [2019], and Sun et al. [2019] assumed the biomarkers trajectories to be monotone functions of age with subject-specific time shifts, adjusting for individual's time of disease onset. Sun et al. [2019]'s work, in particular, statistically resembles a biological neural mass model and assumes that AD pathophysiology influences biomarkers via neuronal activeness, which is quantified by an ensemble of latent binary variables. However, AD pathophysiological process cannot be entirely captured by age with individual shifts; it was first hypothesized in neuropsychology that there exists a latent cognitive process which explains elderly's pathological cognitive decline over time and serves as a shared factor across psychological measurements [Salthouse et al., 1996, Fabrigoule et al., 1998]. Jack Jr et al. [2013] discussed that the ideal way of to present AD biomarker dynamics is along the AD pathophysiological pathway, and that an example would be to represent the disease spectrum using a single latent trait metric, as proposed in Jedynak et al. [2012] and Jedynak et al. [2015]. Jedynak et al. [2012] modeled multiple longitudinal biomarkers as sigmoid functions of a continuous LDP, which is a subject-specific linear function of age. Their estimated temporal ordering of biomarkers largely agrees with Jack Jr et al. [2013]'s, with the exception of one cognitive marker.

We propose a latent variable mixture model, assuming the mixture components to be structural equation models (SEM). Clusters are introduced to account for systematic differences in subjects' disease risk profiles, which reflect subpopulation heterogeneity that may be related to disease resilience or genetic differences and cannot be explained by the biomarkers' average growth curves. According to our method, the subgroup with a higher risk profile is more likely to have worse initial biomarker measurements and to enter a faster deterioration phase in the pathophysiological progression of biomarkers. This is consistent with the findings in Nettiksimmons et al. [2010], in which unsupervised cluster analysis was applied to initial measurements of disease-related biomarkers and a subgroup of CN individuals was identified as having significantly worse cognitive performance at baseline and deteriorating faster over time.

Our proposed model has the following contributions. First, the continuous LDP in the latent structure is a natural caption of the biological AD continuum based on information shared across the multiple repeatedly-measured biomarkers. Since there is no gold standard for AD neuropathologic status during the preclinical stages, the latent metric could be used as a composite score for early AD diagnosis. Additionally, our work extends the LDP specification in Jedynak et al. [2012] to include additional covariates so that their impact on pathological development can be estimated, as well as Guassian process-distributed random effects to account for potential subject-specific time-varying heterogeneity and autocorrelation. Second, to our knowledge, this paper is the first to provide a unified framework to simultaneously identify subgroups by systematic disease risk profiles and discover the temporal order of biomarker changes along a latent metric. Third, our model has the potential to be utilized in clinical trials to identify and recruit asymptomatic individuals who are more likely to develop AD symptoms in the near future. Fourth, using Bayesian estimation, our method naturally quantifies the uncertainty inherent in the finding of the biomarkers' temporal sequence.

The motivating data was longitudinally collected from asymptomatic individuals enrolled in the BIOCARD study (Biomarkers of Cognitive Decline Among Normal Individuals). The average enrollment age was over 50 and more than half of the participants had a first-degree relative with dementia. The observational study records biennial measurements of magnetic resonance imaging (MRI) scans and cerebrospinal fluid (CSF), as well as annual clinical assessments and cognitive testings. The assessments classify individuals as CN, MCI, or dementia. Furthermore, because AD accounts for roughly 70% of dementia cases, the dementia status recorded in the data could be caused by factors other than AD [Albert et al., 2014]; it is considered a clinical diagnostic of AD but may be imprecise. Our joint modeling of the multivariate longitudinal markers from MRI, CSF, and cognitive tests is a variation of the generalized linear mixed effects models (GLMM) [Laird and Ware, 1982], which have been widely used to account for substantial variation in the timing and temporal dynamics of disease progression. Roy and Lin [2000] proposed a latent time joint mixed effects model for multiple continuous outcomes, assuming a linear Gaussian-distributed LDP. Proust et al. [2006] extended Roy and Lin [2000] such that biomarkers are linked to the Gaussian latent process by a family of nonlinear transformations and covariates are included to affect the LDP and the observed outcomes; their work made a connection between the latent time joint mixed effect models and the SEM. Lai et al. [2016] proposed a mixture model clustering method with components being the SEM in Proust et al. [2006].

Section 3 introduces the formulation of our model. Beyond the latent structure and measurement models, we hypothesize that there are different subgroups with systematic shifts in the LDP. Section 4 describes simulation results and demonstrates the validity of the estimation process. Section 5 focuses on the application of our approach to the BIOCARD cohort data, where BICs are derived to determine the number of clusters. Given that the LDP quantifies the AD pathophysiological pathway and that there is no gold standard AD status, we evaluate our method by comparing the rate and duration of conversions into MCI and AD across the sampled posterior clusters. The findings highlight systematic variability in disease progression among individuals and unveil a temporal ordering of the biomarkers that matches the hypotheses in Jack Jr et al. [2013].

3 Method

3.1 The Measurement Model

Suppose the ith subject has J_i time points and K continuous outcomes $\{Y_{ij1}, \ldots, Y_{ijK}\}$ at time $t_{ij}, j = 1, \ldots, J_i$, assuming missing at random among the outcomes. We assume L mixture compo-

nents and quantify subject-specific LDP as the continuous latent scores, $d_i(t) = \{d_i^{(1)}(t), \ldots, d_i^{(L)}(t)\}$. For subject i, the biomarkers are assumed to be independent conditional on $\{\mu_{ik}^{(\ell)}(t); \ell = 1, \ldots, L, k = 1, \ldots, K\}$, which summarizes the observed characteristics and the true disease status represented by the LDP. The distribution of biomarkers can then be written as

$$f(Y_i; \mu_i, \sigma) = \sum_{\ell=1}^{L} \lambda_\ell \left[\prod_{j=1}^{J_i} \prod_{k=1}^{K} f\left(Y_{ijk}; \mu_{ik}^{(\ell)}(t_{ij}), \sigma_k^{(\ell)} \right) \right],$$

where $\lambda = (\lambda_1, \ldots, \lambda_L)$ is the mixing proportions such that $\lambda_\ell \geq 0$ and $\sum_{\ell=1}^{L} \lambda_\ell = 1$. The mixture distribution of the data among N subjects is then $\prod_{i=1}^{N} f(Y_i; \mu_i, \sigma)$. Let U be the subset of biomarkers index where $\{Y_{ijk}; k \in U\}$ are neurological assessments with a ceiling effect, e.g. we observe value u_k when in fact $Y_{ijk} \geq u_k$. This is important because markers such as clinical tests are usually capped by a fixed number of cognitive tasks, imposing a limit to how much cognitive capability a subject can show. When the cognitive performances are assumed to be normal, the biomarkers distribution function can be expressed as a function of the standard gaussian density function ϕ as follows,

$$\begin{aligned} & f\left(Y_{ijk}; \mu_{ik}^{(\ell)}(t_{ij}), \sigma_k^{(\ell)} \right) \\ &= \begin{cases} \phi(\frac{Y_{ijk} - \mu_{ik}^{(\ell)}(t_{ij})}{\sigma_k^{(\ell)}}) & \text{if } k \in U^c \\ \mathbb{1}\{Y_{ijk} = u_k\} \int_{y_{ijk} \geq u_k} \phi(\frac{y_{ijk} - \mu_{ik}^{(\ell)}(t_{ij})}{\sigma_k^{(\ell)}}) dy_{ijk} + \mathbb{1}\{Y_{ijk} < u_k\} \phi(\frac{Y_{ijk} - \mu_{ik}^{(\ell)}(t_{ij})}{\sigma_k^{(\ell)}}) & \text{if } k \in U \end{cases}. \end{aligned}$$

For the population-average progression of the kth biomarker in the ℓth mixture component, $\mu_{ik}^{(\ell)}(t)$, $X_i(t)^T \beta_k$ represents characteristics-adjusted natural neurodegeneration occurs independently of AD pathology, and $h_k(d_i^{(\ell)}(t_{ij}))$ summarizes the effect of a person's cognitive decline caused by AD development, i.e.

$$\mu_{ik}^{(\ell)}(t) = X_i(t)^T \beta_k + h_k(d_i^{(\ell)}(t_{ij})).$$

We postulate the following pathophysiological pattern of a biomarker Y_{ijk} relative to a latent disease score d,

$$h_k(d) = \frac{\gamma_{k1}}{1 + \exp\{-\gamma_{k2}(d - \gamma_{k3})\}}, \qquad \gamma_{k2} > 0,$$

assuming that it is shared across mixture components. The sigmoid function $h_k(d)$ depends on the parameters $(\gamma_{k1}, \gamma_{k2}, \gamma_{k3})$; it has an upper bound of γ_{k1} when $d \to \infty$ and a lower bound of 0 when $d \to -\infty$. Hence, for a person with absolutely no pathological abnormality, the mean of the kth biomarker is left with long-term aging effect not related to AD, e.g. $X_i(t)^T \beta_k$; while for someone with abnormality, it will go from the line $X_i(t)^T \beta_k$, progress as a sigmoid curve, and reach the line $X_i(t)^T \beta_k + \gamma_{k1}$. Hence, γ_{k1} quantifies the farthest possible aggravation in the kth biomarker due to pathological neurodegeneration.

The parameters $\{\gamma_{k3}; k = 1, \ldots, K\}$ are biomarker-specific inflection points in the sigmoid curves and they describe the disease state at which the deterioration in biomarkers is the most apparent. Hence, the proposal anchors underlying disease progression by the temporal order of when each biomarker reaches its maximum dynamic rate relative to the latent disease scores. Estimating inflection points is critical for determining the temporal sequence of biomarkers progression in relation to AD development..

3.2 The Latent Structural Model

The continuous latent score in the ℓth mixture component at time t for individual i is

$$d_i^{(\ell)}(t) = -\alpha_0^{(\ell)} + Z_i(t)^T\alpha + \theta_i^{(\ell)}(t), \qquad t \geq 0, i \in \{1, \ldots, N\}.$$

In each mixture component, the latent score represents the unobserved AD status shared across the K biomarkers. We assume q vector $Z_i(t)$ to be the observed factors that are known to be associated with the underlying AD pathophysiology; $Z_i(t)$ may overlap with the set of covariates in $X_i(t)$.

Intercept $\alpha_0^{(\ell)}$ is the cluster-specific shift in the LDP that defines different systematic subject profiles in AD progression for the ℓth mixture component, where a higher value indicates a lower disease score, and thus a lower risk profile, i.e. less developed towards AD or higher disease resilience. Similar to Sun et al. [2019], we assume the long-term profile parameter to be constant for a relatively short observation time window. For model identifiability and to prevent label switching, we assume $\alpha_0^{(1)} = 0 < \ldots < \alpha_0^{(L)}$, where a larger cluster index indicates a lower risk profile.

To account for subject-level heterogeneity in the temporal dynamics of the LDP, we assume the time-varying random effects $\theta_i^{(\ell)}(t)$ are Gaussian processes, so that potential nonlinearity over time can be captured by kernel functions. The process $\theta_i^{(\ell)}(t)$ concentrates round zero with variation quantified by the exponentiated quadratic covariance function,

$$k(\theta_i^{(\ell)}(t), \theta_i^{(\ell)}(t'); \tau, \rho^{(\ell)}) = \tau^2 \exp\left\{ -\frac{1}{2\rho^{(\ell)2}}(t - t')^2 \right\},$$

where marginal deviation τ is set to one for identifiability, e.g. $\text{var}(\theta_i^{(\ell)}(t)) = 1$ for any $t \geq 0$, and covariance $k(\cdot, \cdot; 1, \rho^{(\ell)})$ equals correlation. Here, the random effects of a subject at any two time points have a stationary positive correlation, which is completely determined by the time difference. A larger length scale, $\rho^{(\ell)}$, indicates a faster decay in the correlations.

3.3 Prior Specification

Parameters in the measurement model are independently assigned weakly informative proper priors, $\beta_k \sim N(0, 100)$ and $\sigma_k \sim \text{Inv-gamma}(0.01, 0.01)$. Choosing priors for the structural model parameters requires more consideration. Given that the age variable is standardized and $\theta_i^{(\ell)}(t) \sim N(0, 1)$, we know that the random effects $\theta_i^{(\ell)}(t)$ and time, i.e. the transformed age, are mostly in $(-2, 2)$. Note that any time difference is between $(-2) - 2 = -4$ and $2 - (-2) = 4$. Write the coefficient for time in the LDP as α_t, we can then view $\theta_i^{(\ell)}(t)/\alpha_t$ as a subject-record-specific time shift in disease progression,

$$\alpha_t t + \theta_i^{(\ell)}(t) = \alpha_t[t + \theta_i^{(\ell)}(t)/\alpha_t].$$

Thus, the time shift is mostly in $(-2/\alpha_t, 2/\alpha_t)$ and this interval should mostly include all possible values of time differences, such as $(-4, 4)$. A reasonable proper prior for α_t is $N(0, 1)$ because it implies that the parameter largely satisfies $|\alpha_t| \leq 2$ and hence $(-4, 4) \subseteq (-2/\alpha_t, 2/\alpha_t)$ with a high probability. Similar reasoning applies to the coefficient of other covariates in the LDP.

In order for the model to be identifiable, $\gamma_{11}, \ldots, \gamma_{K1}$ are pre-specified and calculated from the observed data as the AD-related maximum possible variation in the biomarkers. To suppress degeneration in the sigmoid curves, $\gamma_{12}, \ldots, \gamma_{K2}$ are given weakly informative gamma prior with shape equals 3 and rate equals 1. For inflection points $\gamma_3 = \{\gamma_{13}, \ldots, \gamma_{K3}\}$, we assume hierarchical priors conditional on

the type of measure the kth biomarker belongs to. For instance, MRI related markers simultaneously reflect brain structure and are considered to have similar AD pathophysiology pattern relative to disease progression. Hence, $\gamma_3 = \{\gamma_3^{type}; type \in \{COG, MRI, CSF\}\}$ has priors $\gamma_3^{type} \sim N(\mu^{type}, 1)$ and $\mu^{type} \sim N(0, 2^2)$. This hierarchical prior specification supplies much variability to the sampling of γ_3 because based on the property of Gaussian distribution, the prior spans the range $(-6, 6)$ with high probability, offering more than sufficient coverage to the latent scores upon which the inflection points anchor. Since inflection points are important anchors within the range of latent scores, we further restrict the shift of latent scores across clusters to remain inside the scope of μ^{type} by specifying the prior distribution for the risk profile parameter $\alpha_0^{(\ell)}$, $\ell > 1$, to be a gamma distribution with shape 2 and rate 1.5.

In the latent structural model, when t and t' are $\rho^{(\ell)}$ apart, the correlation between $\theta_i^{(\ell)}(t)$ and $\theta_i^{(\ell)}(t')$ is $\exp(-\frac{1}{2}) \approx 0.61$. Define Δ_i to be maximum time difference among observed time points of the ith subject, e.g. $\Delta_i = \max_{1 \le j \le J_i}(t_{ij}) - \min_{1 \le j \le J_i}(t_{ij})$. We suppress the length scales $\rho^{(\ell)}$ to be no larger than the maximum distance in observed time, $\max_{1 \le i \le N}(\Delta_i)$, so that a correlation higher than 0.61 is unlikely to occur between two time points that are farther apart than $\max_{1 \le i \le N}(\Delta_i)$. Similarly, length scale is no smaller than the minimum distance in observed time, $\min_{1 \le i \le N}(\Delta_i)$. To achieve this, we assume a prior $\rho^{(\ell)} \sim$ inv-gamma(a_ρ, b_ρ) such that $P(\rho^{(\ell)} \in [\min(\Delta_t), \max(\Delta_t)]) = 98\%$, where hyper-parameters a_ρ and b_ρ are calculated based on the data.

4 Simulation Studies

We create 100 data replicates for $n = 300$ subjects, assuming that each subject visits every two years for a total of 20 years, having $N_i = 10$ follow-up visits. Age at the first visit is bootstrap sampled from the BIOCARD dataset's empirical distribution of baseline age. For each subject, we simulate 9 longitudinal biomarkers from a two-component mixture distribution with subject-specific cluster membership. Subject-specific covariates are assumed to be $X_1 \sim$ Binary(0.6) and $X_2 \sim N(0, 1)$; the measurement model covariates are $X = (t, X_1, X_2)$ and latent structural model covariate is time at each visit. Time-varying random effects are simulated from the gaussian process

$$(\theta_{i1}, \ldots, \theta_{i,N_i}) \sim MVN(0, K(t_i)),$$

where $t_i = (t_{i1}, \ldots, t_{i,N_i})$ and $K_{ij}(t_i) = \exp\left\{ -\frac{1}{2\rho^{(\ell)2}}(t_{ij} - t_{ij'})^2 \right\}$.

Table 1 displays the setting of biomarker-specific parameters in the measurement model under which we generated the simulation replicates. The following latent structural model parameters are set for simulating the LDP: length scale $(\rho^{(1)}, \rho^{(2)}) = (1.3, 0.8)$, mixture probabilities $\lambda = (0.35, 0.65)$, risk profile parameter $\alpha_0^{(2)} = 2$, and the coefficient of time in the latent structure, $\alpha_t = 0.4$. Assumed models under the number of mixture components being $L = 1$, 2, and 3 are applied to each of the 100 simulation replicate. For each model estimation, we set an MCMC burn in of 1000, obtain 1000 posterior samples for all parameters, and two MCMC chains are sampled with one computation core for each chain. For each simulation replicate and L, estimation results are summarized based on the chain that meets the following criteria: (1) no divergence in estimated posterior parameters, (2) does not degenerate in estimated parameters, e.g. MCMC struggles to traverse or was trapped on parameter boundaries, and (3) has the larger posterior among the two supplied MCMC chains.

For ease of explanation, we will focus on the parameter γ_{k2} in the following description of evaluation metrics. In the process of applying the model with L components to the rth simulation replicate, there are 1000 posterior samples drawn from the posterior predictive distribution of γ_{k2}, denoted as $\{\gamma^*_{k2rs}; s = 1, \ldots, 1000\}$, where k, r, and s index the outcome, simulation replicate, and posterior draw, respectively. Summarizing over the posterior draws, we derive the posterior mean $\overline{\gamma}^*_{k2r} = \frac{1}{1000}\sum_{s=1}^{1000}\gamma^*_{k2rs}$ and its 95% credible interval U^*_{k2r}. We further aggregate the results over simulation replicates, where Figures 2, 3, and 4 display the posterior mean and 95% credible interval of the bias in posterior mean estimates, $\{|\overline{\gamma}^*_{k2r} - \gamma_{k2}|; r = 1, \ldots, 100\}$, and Figures 5 and 6 show the posterior coverage of the 95% credible intervals, i.e. $\frac{1}{100}\sum_{r=1}^{100}\mathbb{1}\{\gamma_{k2} \in U^*_{k2r}\}$.

We conclude based on the summaries of bias (Figures 2, 3, and 4) and coverage (Figures 5 and 6) that model parameters that are shared across clusters are estimated similarly well so long as the specified number of clusters L in the model is not less than the truth. Meanwhile, cluster-specific parameters, $\{\rho^{(1)}, \rho^{(2)}, \alpha_0, (\sigma^{(1)}_k)^2, (\sigma^{(2)}_k)^2; k = 1, \ldots, 9\}$, need more consideration. Consider the bias summaries of $(\rho^{(1)}, \rho^{(2)})$ in Figure 4, where the color of the error bar and the shape of the point correspond to the model-assumed L and cluster index, respectively. When $L = 1$, there is only one length scale estimate $\widehat{\rho}$, which is simultaneously compared with the simulation truth of $\rho^{(1)}$ and $\rho^{(2)}$; Figure 4 shows that $\widehat{\rho}$ accurately estimates $\rho^{(1)}$ but leaves $\rho^{(2)}$ undiscovered. When $L = 2$, the assumed number of clusters meets the simulation truth. The length scale estimate for each ℓth cluster, $\widehat{\rho}^{(\ell)}$, then unambiguously estimates $\rho^{(\ell)}$ because of the identifiability restrictions. In Figure 4, $\widehat{\rho}^{(\ell)}$ is a good estimate of $\rho^{(\ell)}$ for each ℓ under $L = 2$. When $L = 3$, the model assumed one additional cluster than the simulation truth, and the estimated second and third clusters collapse to identical solutions. The biases of $\widehat{\rho}^{(2)}$ and $\widehat{\rho}^{(3)}$ estimating $\rho^{(1)}$ are high because they are targeting $\rho^{(2)}$ instead of $\rho^{(1)}$, so the biases of $\widehat{\rho}^{(2)}$ and $\widehat{\rho}^{(3)}$ are displayed with half transparency for estimating $\rho^{(1)}$ as shown in the top left panel of Figure 4, while their biases in estimating $\rho^{(2)}$ are highlighted as solid lines displayed in the top right panel. Similar patterns are observed across cluster-specific parameters: when the assumed number of clusters is less than the truth, parameters estimates tend to be biased with unstable posterior coverage; otherwise, estimations are accurate with additional cluster(s) collapsing down to discover one of the two simulated groups.

5 Applications - BIOCARD Cohort

The BIOCARD dataset includes biomarkers derived and collected from MRI, CSF, and cognitive assessment, etc. We jointly studied the 11 biomarkers: CSF biomarkers include beta amyloid (Aβ) 42 over 40 ratio (Aβ_{42}/Aβ_{40}), p-tau$_{181p}$, and total tau (t-tau); MRI biomarkers include entorhinal cortex thickness (ECT), entorhinal cortex volume (ECV), hippocampal volume (HV), medial temporal lobe (MTL) composite score [Pettigrew et al., 2017], and SPARE-AD score (Spatial Pattern of Abnormalities for Recognition of Early AD)[Davatzikos et al., 2009]; cognitive scores include Digit Symbol Substitution Test (DSST) from the Wechsler Adult Intelligence Scale-Revised [Wechsler, 1981], Logical Memory (LM) delayed recall from the Wechsler Memory Scale-Revised, and MMSE (Mini-Mental State Examination) score. MRI biomarkers are computed as an average over the left and right hemisphere. In addition, the volumetric measures are adjusted for intracranial cavity volume by division.

In the present study, we included participants who were assessed to be CN at the enrollment visit. At the time of recruitment, all participants were between the ages of 20 and 86 and had a minimum of twelve years of education. Patients' demographic and medical information was collected at the baseline. Cognitive test scores, entorhinal cortex thickness, and other MRI/CSF measures of interest

had approximately less than 5%, 12%, and 20% missingness, respectively. After the initial visit, most of the follow-up visits happened every 12 months. Table 2 summarizes the baseline risk variables and outcomes for all patients and stratified by the observed conversions of AD diagnoses, i.e. progression from CN to cognitively unaltered, MCI, or AD at the end of follow-up. Besides baseline information, Table 2 also presents the average observed conversion time for those who were assessed to have MCI or AD before data closure. The measurement model includes risk covariates age, gender, presence of ApoE 4 alleles, and education. The risk factors in the latent structure model are age, ApoE 4 alleles, and the interaction between them.

We use the following Bayesian Information Criterion (BIC) approximation [Delattre et al., 2014, Shen and González, 2021] for model selection,

$$BIC = 2\log f(\hat{\psi}|\mathbf{Y}) + \dim(\psi)\log(N),$$

where $\psi = \{\lambda_\ell, \beta_k, \gamma_k, \sigma_k^{(\ell)}, \rho^{(\ell)}, \alpha, \alpha_0^{(\ell)}; \ell = 1, \ldots, L, \text{ and } k = 1, \ldots, K\}$, $\hat{\psi}$ is the maximum a posteriori (MAP) estimate of the parameters, and $\dim(\cdot)$ is the dimension of a vector. Because the prior densities are not uniform, the posterior probability $f(\hat{\psi}|\mathbf{Y})$ is used instead of the likelihood $\mathcal{L}(\hat{\psi}|\mathbf{Y})$ at MAP in the formula. The BIC is calculated to be 56442.14, 55012.24, and 57183.97 for L =1, 2, and 3. Since a lower BIC suggests a better model, we chose probabilistic clustering under $L = 2$ as the final strategy. Under $L = 2$, the estimated parameters in the measurement model are displayed in Table 5; posterior samples of latent structural model parameters and mixing proportions are summarized in Table 6; and the estimated biomarker-pathophysiological patterns are visualized in Figure 1.

We investigate the dynamic patterns of AD clinical diagnoses, especially subjects' status conversion into MCI or AD, stratified by the risk profiles estimated from the proposed L-group probabilistic clustering. The diagnoses are not used as labels for supervised learning as they might be imprecise and unstable during transition periods. However, diagnostic-based time of status conversion is of essential scientific relevance, so we utilize it as a metric for validating and comparing cluster findings. Let a subject's initial and final diagnostic statuses be determined by the first and last available clinical assessments, and that the conversion time to a non-CN final status is determined by the time between the baseline and when a matching diagnostic first occurred during a subject's follow-up.

Under the selected probabilistic clustering model ($L = 2$), Table 3 summarizes subjects' final diagnostics by posterior sampled cluster membership. We observe that subjects in the second group, i.e. the cluster assessed to have systematically lower risk profile, have a smaller ratio of developing MCI or AD, significantly less cases of AD than MCI, and longer duration in acquiring cognitive impairment. Similar to Table 3, Table 4 provides posterior cluster membership-stratified summaries of subjects' pattern of developing a more severe disease status, e.g. going from symptom-free to MCI and from MCI to AD. Results show that subjects in the lower risk profile group generally have a lower rate and longer duration of disease aggravation.

Specifically, we examine the estimated biomarker-pathophysiological patterns of the 11 measures of interest. The estimation results under $L = 2$ are consistent with relevant scientific evidence on the temporal ordering of biomarkers described in Jack Jr et al. [2013]: first, Aβ_{42}/Aβ_{40} and t-tau are the earliest major AD biomarkers to become aberrant, followed by structural MRI and clinical symptoms; second, Aβ_{42}/Aβ_{40} was completely abnormal years before diagnosis of dementia; third, p-tau$_{181p}$ and t-tau progresses similarly over time, becoming increasingly abnormal as the disease progresses; and fourth, Aβ_{42}/Aβ_{40} is more abnormal early in the course of the disease than hippocampus volume.

6 Discussion

Using the BIOCARD data, this study proposed a novel unsupervised machine learning approach for (1) generating an AD risk score among initially CN individuals to describe the progression of AD, (2) identifying different groups of subjects with systematically distinct risk profiles of developing AD symptoms, and (3) discovering the temporal sequence of multiple biomarkers as AD progresses. In terms of external validation, clusters identified based on latent risk profiles well separates diagnostic-based status conversions into MCI or AD. The results show that the AD risk score, the LDP from the unsupervised machine learning, captures the underlying disease burden well and can potentially be applied to individuals in other stages of the AD continuum.

The proposed method has several advantages. The probabilistic clustering approach allows for the early detection of asymptomatic subjects who have a systematically higher risk profile of developing AD; this has the potential to significantly benefit AD-related clinical trial designs, allowing studies to more efficiently recruit asymptomatic subjects who have a higher chance of developing AD in the near future. A byproduct of the approach is the discovery of the temporal order in which biomarkers change in relation to AD progression, and the conclusions are consistent with medical hypotheses and findings. Furthermore, the latent variable framework provides greater flexibility and capability in accounting for more aspects of the disease mechanism, allowing for a more robust description of the underlying pathophysiology. The method can also be easily modified to include new biomarkers such as medical co-morbidities and genetic sequencing. The Bayesian approach offers additional benefits by quantifying stochastic uncertainty in the estimation of cluster membership and the temporal order of biomarkers by AD progression.

One limitation of this study is that the clustering divides the population into groups with relatively homogeneous AD progression patterns. Our method does not identify clusters based on a gold standard disease outcome or on AD clinical diagnoses, which are imprecise surrogates for true AD status. The estimated LDP is considered to reflect AD pathophysiology because biomarkers are selected to be biologically relevant and, as with AD pathology, the LDP is the essential shared factor that contributes to the heterogeneity of biomarkers. Furthermore, the method is limited to subjects who were initially asymptomatic; the LDP lacks an anchor that allows application on data from subjects who were cognitively abnormal at enrollment.

Possible extensions of the proposal include (1) use semi-parametric transformation of the outcomes with link functions to allow for non-Gaussian biomarkers of various types; (2) specify a threshold parameter in the LDP as an anchor on clinical diagnoses of cognitive abnormality; (3) jointly model the longitudinal biomarkers with competing risks events such as depression and death; and (4) allow covariates to also contribute to mixing proportions and relax the assumption of linear trends in covariates.

References

M. Albert, A. Soldan, R. Gottesman, G. McKhann, N. Sacktor, L. Farrington, M. Grega, R. Turner, Y. Lu, S. Li, et al. Cognitive changes preceding clinical symptom onset of mild cognitive impairment and relationship to apoe genotype. *Current Alzheimer Research*, 11(8):773–784, 2014.

A. Association et al. 2022 alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 18(4): 700–789, 2022.

D. A. Bennett, R. S. Wilson, J. A. Schneider, D. A. Evans, L. A. Beckett, N. T. Aggarwal, L. L. Barnes, J. H. Fox, and J. Bach. Natural history of mild cognitive impairment in older persons. *Neurology*, 59(2):198–205, 2002.

C. Davatzikos, F. Xu, Y. An, Y. Fan, and S. M. Resnick. Longitudinal progression of alzheimer's-like patterns of atrophy in normal older adults: the spare-ad index. *Brain*, 132(8):2026–2035, 2009.

M. Delattre, M. Lavielle, and M.-A. Poursat. A note on bic in mixed-effects models. *Electronic journal of statistics*, 8(1):456–475, 2014.

M. C. Donohue, H. Jacqmin-Gadda, M. Le Goff, R. G. Thomas, R. Raman, A. C. Gamst, L. A. Beckett, C. R. Jack Jr, M. W. Weiner, J.-F. Dartigues, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10:S400–S410, 2014.

B. Dubois, H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, et al. Preclinical alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3):292–323, 2016.

C. Fabrigoule, I. Rouch, A. Taberly, L. Letenneur, D. Commenges, J.-M. Mazaux, J.-M. Orgogozo, and J.-F. Dartigues. Cognitive process in preclinical phase of dementia. *Brain: a journal of neurology*, 121(1):135–141, 1998.

H. Hampel, K. Burger, S. J. Teipel, A. L. W. Bokde, H. Zetterberg, and K. Blennow. Core candidate neurochemical and imaging biomarkers of alzheimer's disease. *Alzheimer's & Dementia*, 4(1):38–48, 2008.

S. D. Han, J. Gruhl, L. Beckett, H. H. Dodge, N. H. Stricker, S. Farias, and D. Mungas. Beta amyloid, tau, neuroimaging, and cognition: sequence modeling of biomarkers for alzheimer's disease. *Brain imaging and behavior*, 6(4):610–620, 2012.

C. R. Jack Jr, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski. Hypothetical model of dynamic biomarkers of the alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

C. R. Jack Jr, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, et al. Update on hypothetical model of alzheimer's disease biomarkers. *Lancet neurology*, 12(2):207, 2013.

B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, J. L. Prince, et al. A computational neurodegenerative disease progression score: method and results with the alzheimer's disease neuroimaging initiative cohort. *Neuroimage*, 63(3):1478–1486, 2012.

B. M. Jedynak, B. Liu, A. Lang, Y. Gel, J. L. Prince, A. D. N. Initiative, et al. A computational method for computing an alzheimer's disease progression score; experiments and validation with the adni data set. *Neurobiology of aging*, 36:S178–S184, 2015.

D. Lai, H. Xu, D. Koller, T. Foroud, and S. Gao. A multivariate finite mixture latent trajectory model with application to dementia studies. *Journal of applied statistics*, 43(14):2503–2523, 2016.

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

D. Li, S. Iddi, W. K. Thompson, M. C. Donohue, and A. D. N. Initiative. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Statistical methods in medical research*, 28 (3):835–845, 2019.

J. Nettiksimmons, D. Harvey, J. Brewer, O. Carmichael, C. DeCarli, C. R. Jack Jr, R. Petersen, L. M. Shaw, J. Q. Trojanowski, M. W. Weiner, et al. Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. *Neurobiology of aging*, 31 (8):1419–1428, 2010.

R. C. Petersen, R. O. Roberts, D. S. Knopman, B. F. Boeve, Y. E. Geda, R. J. Ivnik, G. E. Smith, and C. R. Jack. Mild cognitive impairment: ten years later. *Archives of neurology*, 66(12):1447–1455, 2009.

C. Pettigrew, A. Soldan, K. Sloane, Q. Cai, J. Wang, M.-C. Wang, A. Moghekar, M. I. Miller, M. Albert, B. R. Team, et al. Progressive medial temporal lobe atrophy during preclinical alzheimer's disease. *NeuroImage: Clinical*, 16:439–446, 2017.

C. Proust, H. Jacqmin-Gadda, J. M. Taylor, J. Ganiayre, and D. Commenges. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*, 62(4): 1014–1024, 2006.

J. Roy and X. Lin. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics*, 56(4):1047–1054, 2000.

T. A. Salthouse, H. E. Hancock, E. J. Meinz, and D. Z. Hambrick. Interrelations of age, visual acuity, and cognitive functioning. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 51(6):P317–330, 1996.

N. Shen and B. González. Bayesian information criterion for linear mixed-effects models. *arXiv preprint arXiv:2104.14725*, 2021.

R. A. Sperling, P. S. Aisen, L. A. Beckett, D. A. Bennett, S. Craft, A. M. Fagan, T. Iwatsubo, C. R. Jack Jr, J. Kaye, T. J. Montine, et al. Toward defining the preclinical stages of alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):280–292, 2011.

M. Sun, D. Zeng, and Y. Wang. Leveraging nonlinear dynamic models to predict progression of neuroimaging biomarkers. *Biometrics*, 75(4):1240–1252, 2019.

L. Vermunt, S. A. Sikkes, A. Van Den Hout, R. Handels, I. Bos, W. M. Van Der Flier, S. Kern, P.-J. Ousset, P. Maruff, I. Skoog, et al. Duration of preclinical, prodromal, and dementia stages of alzheimer's disease in relation to age, sex, and apoe genotype. *Alzheimer's & Dementia*, 15(7): 888–898, 2019.

Z. Wang and X.-H. Zhou. Biomarker assessment and combination with differential covariate effects and an unknown gold standard, with an application to alzheimer's disease. *The Annals of Applied Statistics*, 12(2):1204–1227, 2018.

Z. Wang, Z. Tang, Y. Zhu, C. Pettigrew, A. Soldan, A. Gross, and M. Albert. Ad risk score for the early phases of disease based on unsupervised machine learning. *Alzheimer's & Dementia*, 16(11): 1524–1533, 2020.

D. Wechsler. Wechsler adult intelligencescale-revised. *New York: The psychological corporation*, 1981.

Tables

	Outcome Index k								
	1	2	3	4	5	6	7	8	9
β_{k0}	0.2	0.1	0	-0.3	-0.4	-0.3	-1	-1	-0.1
β_{kt}	-0.1	0.2	0.2	0.1	0.2	0.4	0.1	0.1	0.1
β_{kX_1}	-0.3	-0.4	0	0.2	0.4	0.3	0	0	0
β_{kX_2}	-0.1	-0.2	0.1	0	0	0	0	0	0
γ_{k1}	2	2	1	1	1	2	2	2	1
γ_{k2}	3	3	4	4	4	4	2	2	4
γ_{k3}	1	0.8	-0.2	-0.2	0	0.2	-2	-2	-1
σ^2_{k1}	0.8	0.7	0.9	0.8	0.6	0.3	1	1	1
σ^2_{k2}	0.8	0.7	0.6	0.8	0.6	0.6	0.04	0.04	0.5

Table 1: Values of the measurement model parameters for generating simulation datasets.

	All (N=297)	NM-NM (N=215)	NM-MCI (N=53)	NM-AD (N=29)
Conversion Time			12.61 (4.25, 23.12)	13.45 (3.01, 19.17)
Age	57.21 (31.6, 76.76)	55.45 (28.87, 73.44)	60.06 (36.99, 73.61)	65.1 (38.55, 80.18)
Apoe	0.61	0.60	0.57	0.69
Gender	0.59	0.60	0.62	0.41
Education	-0.07 (-2.23, 1.2)	-0.02 (-2.23, 1.2)	-0.3 (-2.23, 1.2)	-0.06 (-2.23, 1.2)
MMSE Score	-0.27 (-0.59, 1.54)	-0.3 (-0.59, 0.83)	-0.17 (-0.59, 1.54)	-0.18 (-0.59, 0.83)
LM	0.69 (-1.12, 2.77)	0.57 (-1.12, 2.53)	0.99 (-0.63, 2.77)	1.02 (-1.12, 2.53)
DSST	0.14 (-1.7, 1.86)	-0.04 (-1.86, 1.86)	0.56 (-0.91, 1.94)	0.62 (-0.43, 1.55)
ECT	-0.02 (-1.58, 2.04)	-0.11 (-1.74, 1.74)	0.09 (-1.49, 2.04)	0.48 (-1.11, 1.96)
HV	-0.16 (-1.77, 1.66)	-0.22 (-1.84, 1.43)	-0.16 (-1.67, 1.56)	0.28 (-1.45, 1.49)
ECV	-0.04 (-2.07, 1.56)	-0.04 (-1.44, 1.45)	-0.08 (-2.77, 1.48)	0.06 (-1.74, 1.69)
MTL	-0.16 (-2.08, 1.79)	-0.19 (-1.93, 1.43)	-0.17 (-2.24, 1.84)	0.11 (-2.12, 2.31)
SPARE-AD	-0.25 (-1.8, 1.47)	-0.31 (-1.79, 1.28)	-0.18 (-1.8, 0.95)	0.09 (-1.3, 1.57)
T-tau	-0.23 (-1.18, 2.17)	-0.37 (-1.25, 0.78)	-0.05 (-0.95, 1.74)	0.62 (-1, 2.69)
P-tau$_{181p}$	-0.24 (-1.16, 2.27)	-0.39 (-1.17, 0.98)	-0.06 (-1.05, 2.24)	0.6 (-0.89, 2.67)
Aβ_{42}/Aβ_{40}	-0.12 (-1.23, 1.79)	-0.28 (-1.33, 1.64)	0.21 (-0.89, 1.68)	0.56 (-0.92, 1.74)

Table 2: Summary of covariates and biomarkers at the baseline for all subjects and three subgroups of those who stayed normal, obtained MCI, and AD at the time of data closure.

Cluster Index	Subject Count	Last Diagnosis	Percentage	Duration
1	100.89 (95, 107)	CN	0.59 (0.57, 0.62)	
		MCI	0.2 (0.19, 0.22)	11.35 (11.2, 11.65)
		AD	0.2 (0.19, 0.21)	12.26 (12.22, 12.33)
2	196.11 (190, 202)	CN	0.79 (0.78, 0.8)	
		MCI	0.17 (0.16, 0.17)	13.41 (13.14, 13.57)
		AD	0.04 (0.04, 0.05)	16.21 (15.64, 16.37)

Table 3: Summary of the last available diagnosis for subjects in each cluster from the probabilistic clustering model under $L = 2$. The table shows posterior mean and 95% credible interval of the percentage of subjects in the cluster with the last diagnosis type, as well as the time from enrollment to the first occurrence of a diagnosis that matches the last diagnosis.

Cluster Index	Conversion Type	Percentage	Duration
1	CN-MCI	0.38 (0.36, 0.41)	10.54 (10.44, 10.7)
	MCI-AD	0.45 (0.42, 0.47)	3.27 (3.23, 3.42)
2	CN-MCI	0.27 (0.26, 0.28)	12.04 (11.9, 12.13)
	MCI-AD	0.15 (0.13, 0.16)	3.77 (3.43, 3.91)

Table 4: Summary of conversion from the first diagnosis of a state A to the first diagnosis of a state B for subjects in each cluster from the probabilistic clustering model under $L = 2$, where states $(\text{A, B}) \in \{(\text{CN, MCI}), (\text{MCI, AD})\}$. Posterior mean and 95% credible interval of the percentage and duration of conversion are calculated. Percentage is defined as the ratio of subjects who developed state B among those who were in state A.

	β_0	β_t	β_{apoe}	β_{sex}	$\beta_{education}$
MMSE Score	-0.77 (-0.9, -0.65)	0.45 (0.4, 0.51)	-0.01 (-0.11, 0.11)	-0.24 (-0.36, -0.13)	-0.2 (-0.26, -0.15)
LM	0.18 (0.11, 0.25)	-0.1 (-0.13, -0.07)	-0.11 (-0.18, -0.04)	-0.36 (-0.42, -0.29)	-0.11 (-0.14, -0.08)
DSST	0.14 (0.08, 0.21)	0.24 (0.21, 0.28)	-0.11 (-0.18, -0.04)	-0.39 (-0.45, -0.33)	-0.19 (-0.22, -0.15)
ECT	-0.13 (-0.25, -0.01)	0.09 (0.03, 0.15)	-0.21 (-0.3, -0.1)	0.35 (0.24, 0.46)	0.1 (0.04, 0.16)
HV	0.22 (0.14, 0.31)	0.17 (0.13, 0.21)	-0.19 (-0.27, -0.11)	-0.4 (-0.48, -0.32)	0.08 (0.04, 0.13)
ECV	-0.27 (-0.36, -0.18)	0.1 (0.06, 0.14)	0.06 (-0.02, 0.15)	0.16 (0.07, 0.25)	-0.03 (-0.08, 0.01)
MTL	-0.43 (-0.51, -0.35)	0.22 (0.18, 0.26)	-0.06 (-0.14, 0.02)	0.44 (0.36, 0.52)	0.05 (0.01, 0.09)
SPARE-AD	-0.33 (-0.41, -0.26)	0.38 (0.33, 0.41)	-0.09 (-0.16, -0.02)	0.26 (0.18, 0.33)	0.07 (0.03, 0.1)
T-tau	-1 (-1.04, -0.96)	0.1 (0.08, 0.13)	-0.05 (-0.09, -0.02)	0.02 (-0.03, 0.06)	-0.04 (-0.06, -0.01)
P-tau$_{181p}$	-1.11 (-1.14, -1.07)	0.11 (0.09, 0.13)	-0.11 (-0.15, -0.07)	0 (-0.04, 0.04)	-0.02 (-0.04, 0.01)
Aβ_{42}/Aβ_{40}	-0.12 (-0.2, -0.05)	0.13 (0.09, 0.17)	-0.24 (-0.32, -0.18)	-0.03 (-0.11, 0.04)	0.01 (-0.02, 0.05)
	$(\sigma^{(1)})^2$	$(\sigma^{(2)})^2$	γ_2	γ_3	
MMSE Score	2.02 (1.77, 2.31)	2.13 (1.91, 2.4)	7.36 (4.44, 11.71)	1.35 (1.09, 1.63)	
LM	0.76 (0.67, 0.86)	0.79 (0.73, 0.85)	10.45 (6.6, 15.15)	0.96 (0.74, 1.21)	
DSST	0.69 (0.61, 0.78)	0.75 (0.7, 0.79)	9.65 (5.32, 15.14)	0.77 (0.55, 1.03)	
ECT	0.73 (0.64, 0.82)	0.97 (0.87, 1.08)	6.2 (3.89, 10.04)	0.83 (0.56, 1.08)	
HV	0.9 (0.8, 1.02)	0.62 (0.56, 0.68)	10.18 (6.57, 15.74)	-0.15 (-0.42, 0.11)	
ECV	0.86 (0.76, 0.97)	0.8 (0.73, 0.86)	9.96 (6.42, 14.95)	-0.11 (-0.36, 0.12)	
MTL	0.65 (0.57, 0.73)	0.59 (0.54, 0.64)	11.66 (8.63, 15.05)	-0.05 (-0.3, 0.18)	
SPARE-AD	0.34 (0.29, 0.39)	0.56 (0.51, 0.6)	11.36 (8.29, 15.15)	0.2 (-0.01, 0.42)	
T-tau	1.53 (1.34, 1.75)	0.03 (0.03, 0.04)	1.69 (1.52, 1.86)	-2.21 (-2.51, -1.93)	
P-tau$_{181p}$	1.42 (1.28, 1.58)	0.01 (0.01, 0.01)	1.31 (1.2, 1.43)	-2.14 (-2.44, -1.85)	
Aβ_{42}/Aβ_{40}	1.22 (1.06, 1.42)	0.47 (0.43, 0.51)	9.09 (4.61, 16.19)	-1.16 (-1.44, -0.89)	

Table 5: Summary of estimated posterior mean and 95% credible interval of the biomarker-specific parameters in the measurement model under $L = 2$.

$\alpha_0^{(2)}$	α_{apoe}	α_{age}	$\alpha_{apoe\times age}$
2.38 (2.09, 2.70)	0 (-0.19, 0.20)	0.47 (0.33, 0.61)	-0.18 (-0.33, -0.03)
$\rho^{(1)}$	$\rho^{(2)}$	$\lambda^{(1)}$	$\lambda^{(2)}$
1.50 (1.20, 1.87)	0.85 (0.77, 0.94)	0.35 (0.28, 0.42)	0.65 (0.58, 0.72)

Table 6: Summary of estimated posterior mean and 95% credible interval of the parameters in the latent structural model and the mixing component under $L = 2$.

Figures

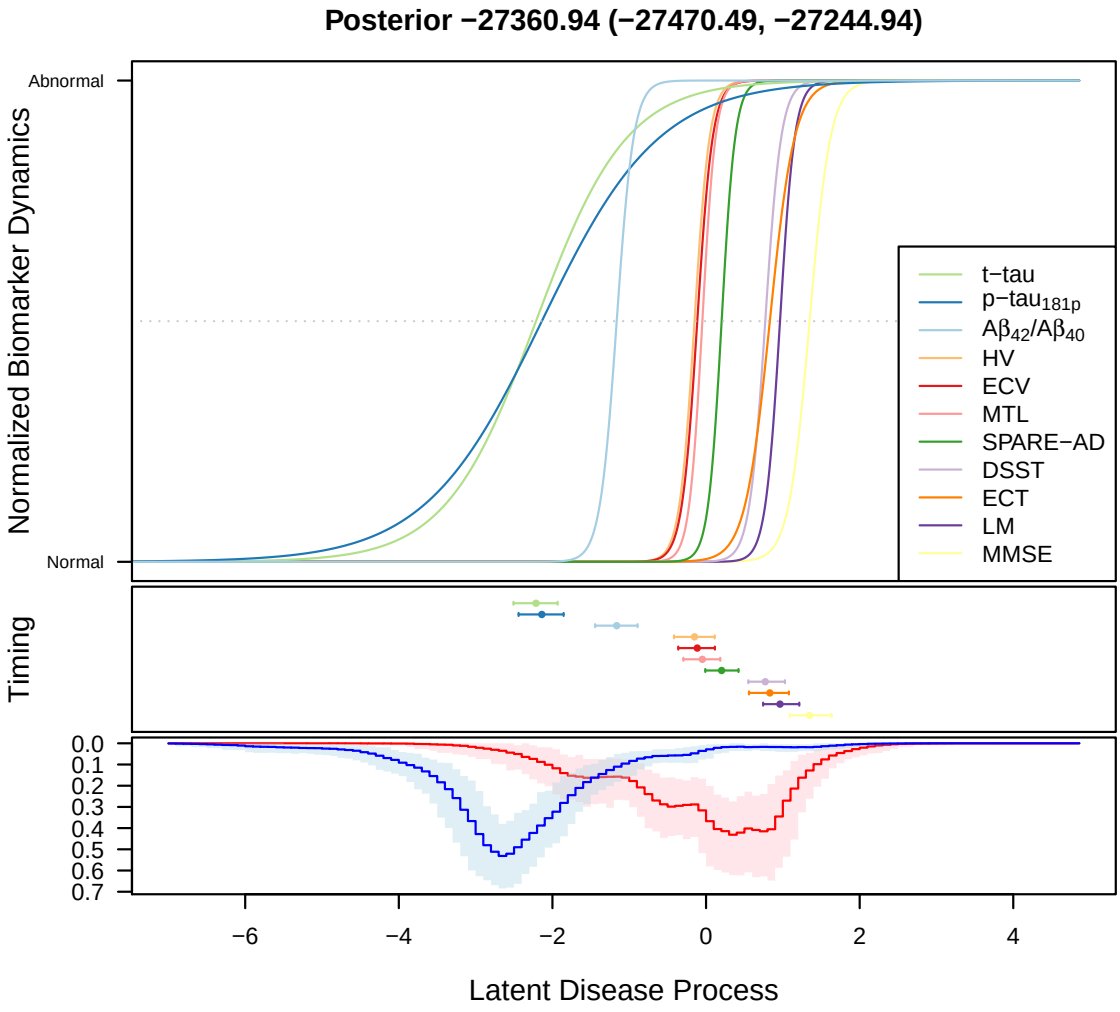

Figure 1: Under $L = 2$, (top figure) estimated $h_k(d)$ as a function of latent score d for the 11 biomarkers indexed by k (assuming the same start and end-level among biomarkers), where green, red, and black correspond to CSF, MRI, and cognitive assessments, respectively; and (bottom figure) the estimated distribution of $\{d_i^{(\ell)}(t_{ij}); Z_i = \ell, j = 1, \ldots, J_i, \text{ and } i = 1, \ldots, N\}$, where blue $(\ell = 1)$ are those more progressed in AD and red $(\ell = 2)$ are those with higher disease resilience. Solid lines represent the posterior mean estimates and bands reflect the 95% credible intervals.

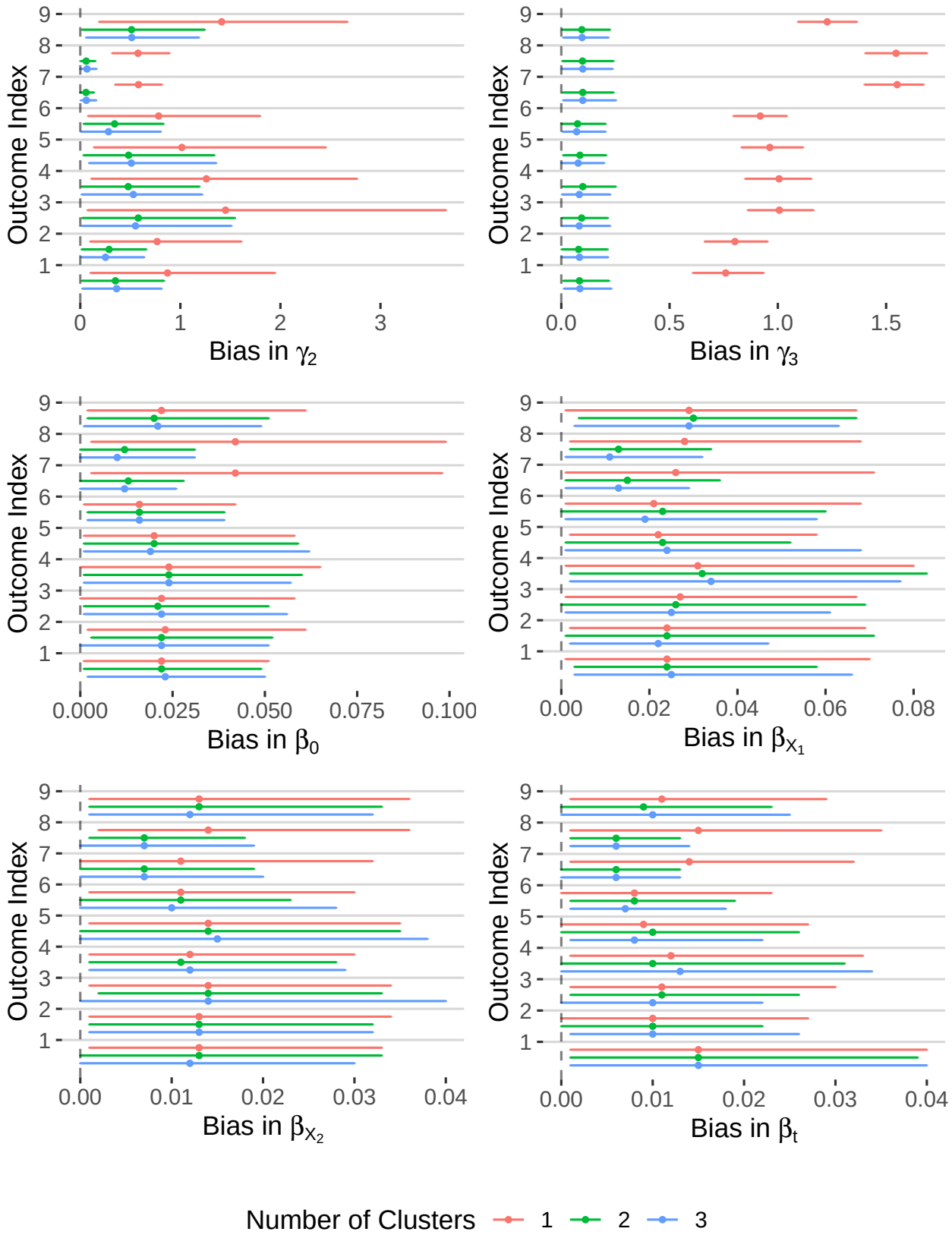

Figure 2: Simulation MSE from the estimation of pathophysiological parameters (γ_2, γ_3) and natural neurodegeneration parameters $(\beta_0, \beta_{X_1}, \beta_{X_2}, \beta_t)$.

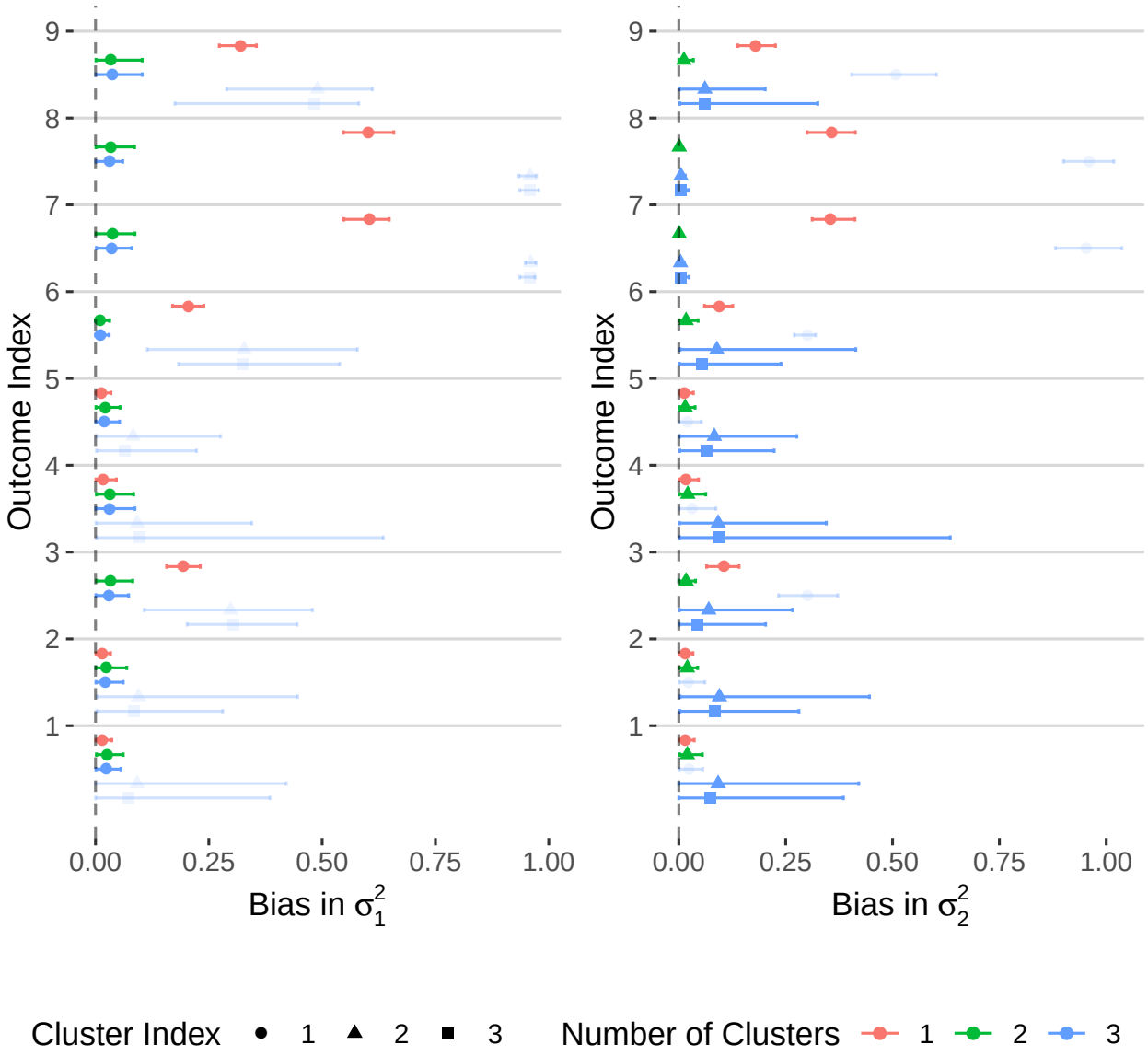

Figure 3: Simulation MSE from the estimation of σ_1^2 and σ_2^2.

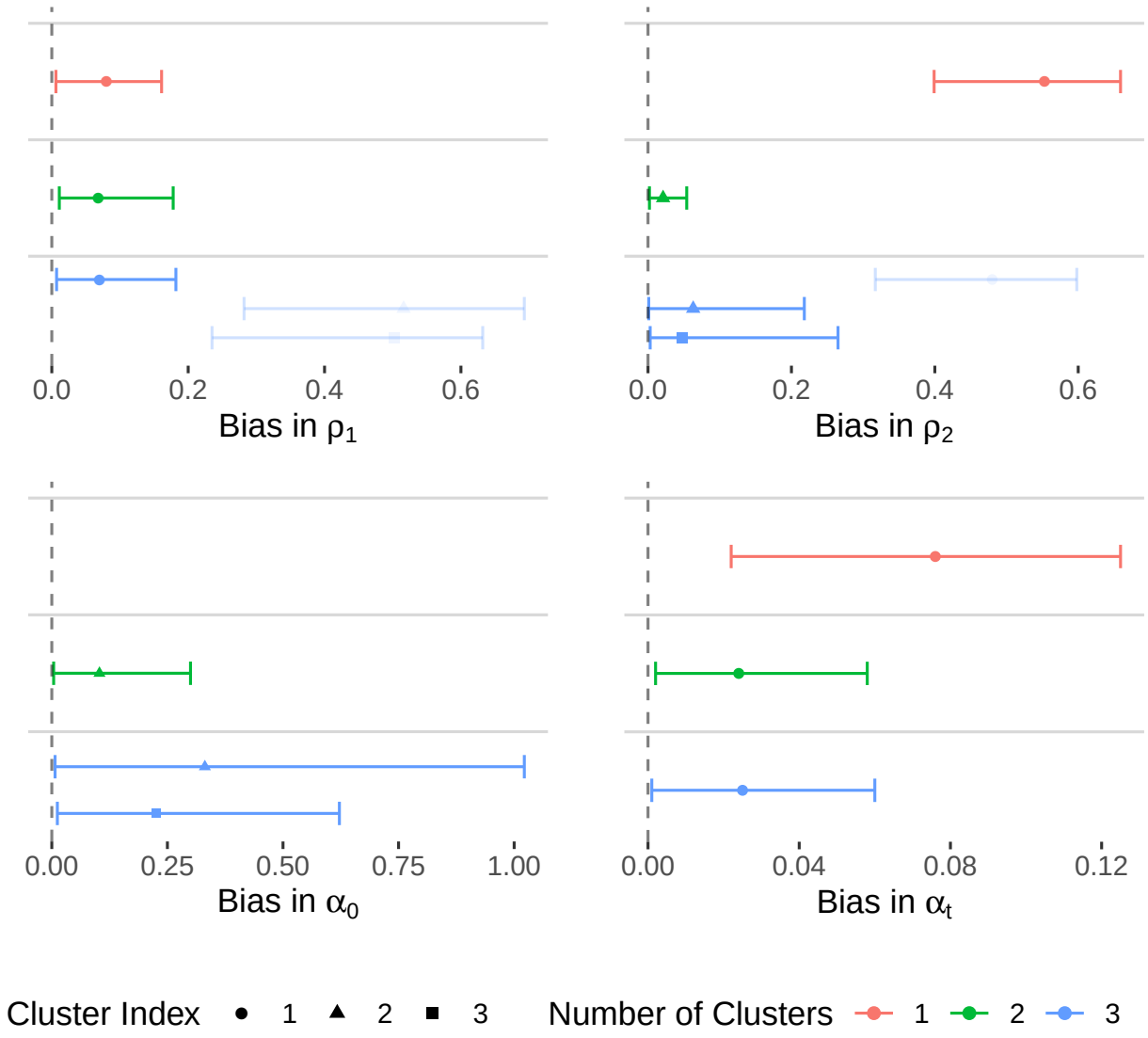

Figure 4: Simulation MSE from the estimation of autoregressive random effects' length scale (ρ_1, ρ_2) and LDP coefficients (α_0, α_t).

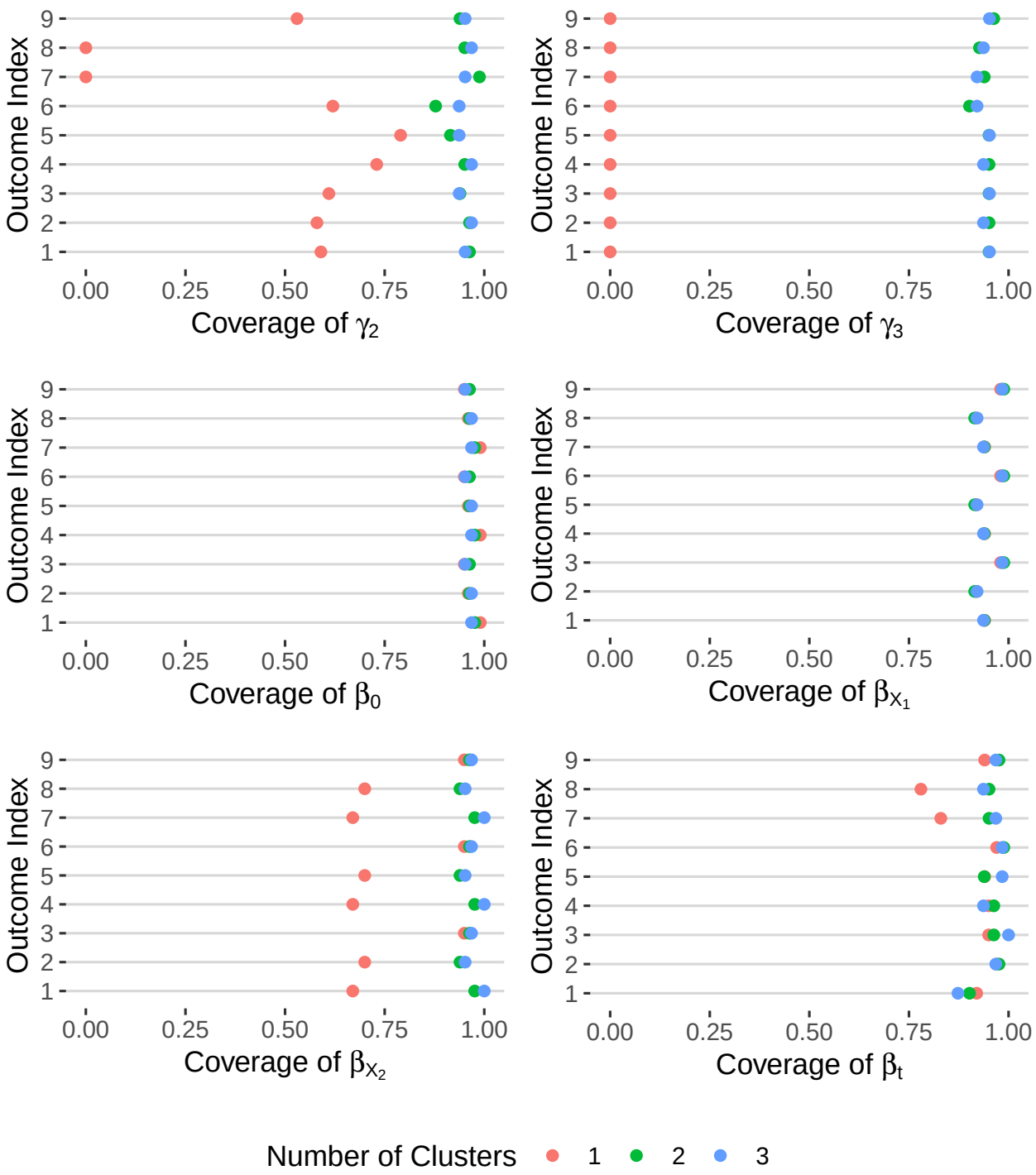

Figure 5: Simulation posterior coverage from the estimation of (γ_2, γ_3) and $(\beta_0, \beta_{X_1}, \beta_{X_2}, \beta_t)$.

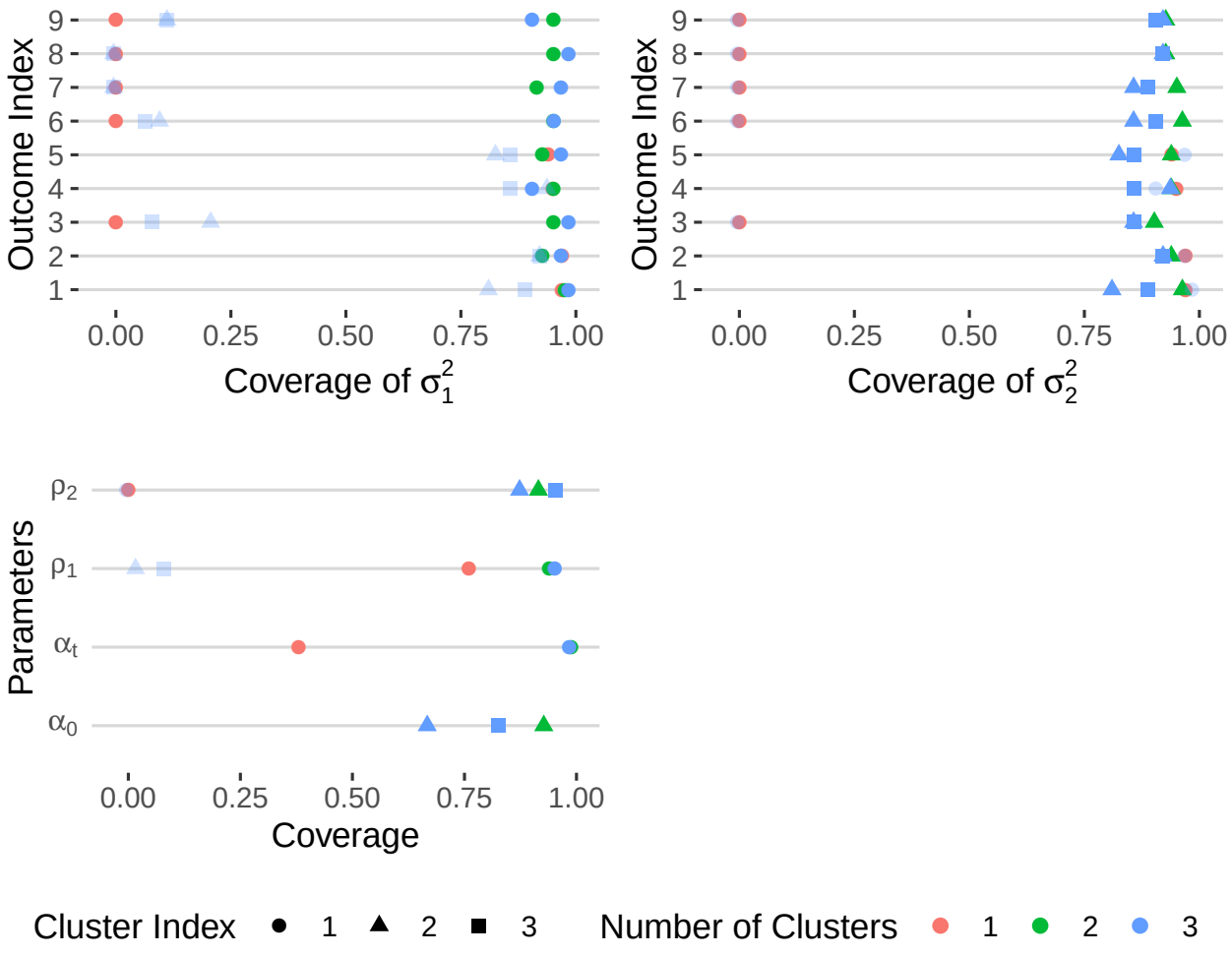

Figure 6: Simulation posterior coverage from the estimation of (σ_1^2, σ_2^2), (ρ_1, ρ_2), and (α_0, α_t).