Stage-Aware Event-Based Modeling (SA-EBM) for Disease Progression

Hongtao Hao HONGTAOH@CS.WISC.EDU
University of Wisconsin-Madison

Vivek Prabhakaran VPRABHAKARAN@UWHEALTH.ORG
University of Wisconsin-Madison

Veena A Nair VNAIR@UWHEALTH.ORG
University of Wisconsin-Madison

Nagesh Adluru ADLURU@WISC.EDU
University of Wisconsin-Madison

Joseph L. Austerweil JOSEPH.AUSTERWEIL@GMAIL.COM
Chiba Institute of Technology, & University of Wisconsin-Madison

for the Alzheimer's Disease Neuroimaging Initiative*

Abstract

As diseases progress, they increasingly impact more cognitive and biological factors. By formulating probabilistic models with this basic assumption, Event-Based Models (EBMs) enable researchers to discover the progression of a disease that makes earlier diagnosis and effective clinical interventions possible. We build on prior EBMs with two major improvements: (1) dynamic estimation of healthy and pathological biomarker distributions, and (2) explicit modeling of disease stage distribution. We tested existing approaches and our novel approach on 9,000 synthetic datasets and also the real-world ADNI data. We found that our stage-aware EBM (SA-EBM) significantly outperforms prior methods, such as Gaussian Mixture Model (GMM) EBM, Kernel Density Estimation EBM and Discriminative EBM, in accurately recovering the order of disease events and assigning individual disease stages. Our package can be installed by `pip install pysaebm`. Source codes for the package, experiments, and visualizations are available in Appendix N, or at https://saebm.hongtaoh.com.

1. Introduction

Understanding how diseases progress over time is central to early diagnosis, prognosis, and intervention. This is especially the case for chronic and neurodegenerative conditions such as Alzheimer's and related dementias (ADRDs) including post-stroke vascular contributions to cognitive impairments and dementia (VCID) and frontotemporal lobar dementia (FTLD), and Parkinson's disease. While longitudinal studies are ideal, they are often expensive, time-

*. 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 in Appendix O.

consuming, and logistically challenging (Young et al., 2024), resulting in limited availability. As a result, there is increasing interest in inferring disease progression from cross-sectional data, where there is a single data point per participant.

Table 1 represents a typical cross-sectional dataset containing biomarker measurements from both healthy and progressing participants. The challenge is to infer the temporal sequence where biomarkers become pathological as the disease develops. This table clearly illustrates that the task is daunting without the support of advanced statistical models.

Table 1: Participant biomarker measurements

ID	Impacted	FUS-FCI	P-Tau	MMSE	AB	HIP-FCI	PCC-FCI	⋯
1	Yes	27.16	-6.14	24.49	147.99	1.59	2.80	⋯
2	Yes	17.20	57.89	24.43	157.13	-4.06	8.84	⋯
3	Yes	13.99	62.51	20.87	158.12	6.48	4.42	⋯
4	No	3.38	26.23	27.05	275.64	-2.94	10.68	⋯
5	No	9.90	20.60	28.97	242.11	-2.81	5.69	⋯
6	No	9.29	40.00	26.53	343.85	-3.56	7.31	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱

Note: Refer to Table 2 in Appendix A for more detailed information about biomarkers.

To enable accurate inferences of the progression order from cross-sectional datasets, the Event-Based Model (EBM; Fonteijn et al., 2012) posits that each biomarker from an impacted participant is associated with an event that encodes whether the biomarker is affected by the disease and thus generated from an atypical distribution. The disease follows a latent progression across biomarkers, with each participant occupying an unknown stage along this trajectory. The EBM uncovers this trajectory by analyzing biomarker patterns across participants, even using only cross-sectional observations.

The EBM has been applied to cross-sectional participant data of a number of different diseases (Chen et al., 2016; Eshaghi et al., 2018; Hu et al., 2025). Despite the remarkable progress of different EBMs over the last decade, there are still limitations in existing approaches. For example, in the calculation of data likelihood, there is typically an implicit assumption of a uniform distribution over disease stages; However, participants in severe stages may be underrepresented in clinical studies (Donohue et al., 2014), resulting in non-uniform empirical distributions. Further, the estimation of healthy and atypical biomarker distributions is usually done without conducting inference about the distribution of disease stages and about the progression order. This can result in less accurate estimates. Thus, for the most accurate inferences, the disease progression order and biomarker distributions should be estimated in a joint manner.

In this work, we introduced Stage-Aware Event Based Modeling (SA-EBM), which addresses these two challenges by jointly determining progression order, disease stage distribution, and biomarker distributions. We evaluated SA-EBM on a comprehensive set of 9,000 synthetic datasets. It demonstrated improved performance compared to the state-of-the-art EBM methods in both ordering (i.e., recovering the order of disease events) and staging (i.e., assigning individual disease stages) tasks. Our results highlighted the robustness of SA-EBM across various progression scenarios, including varied disease stage distributions, disease progression simulation frameworks, and biomarker distributions that may deviate

from SA-EBM's assumptions. We also applied the method to ADNI (Mueller et al., 2005), a large data set of patients with neurodegenerative diseases and matched controls. We found that the SA-EBM estimated participant stages consistent with the stage of their clinical diagnosis (unobserved by the model) and a disease progression ordering that is partially consistent with the current scientific consensus.

In summary, our contributions are:

1. We propose a novel Stage-Aware EBM framework that jointly models and updates the disease stage distribution, the biomarker progression order, and the parameters of biomarker distributions.

2. We implement five parameter estimation methods within the proposed SA-EBM framework.

3. We benchmark our methods and existing EBM algorithms on 9,000 synthetic datasets inspired by hypothetical and real-world clinical data, and also on the real-world ADNI dataset (Mueller et al., 2005). We explored the effect of sample sizes and proportions of healthy (control) samples on model performance.

4. We demonstrate that SA-EBM methods achieve robust improvements in ordering and staging accuracy compared to benchmark algorithms.

Generalizable Insights about Machine Learning in the Context of Healthcare

Our work provides several insights applicable to machine learning in healthcare beyond disease progression modeling:

- Simpler models might outperform sophisticated ones: Our Gaussian assumption-based models consistently outperform complicated KDE-based approaches, even in data with irregular patterns, demonstrating that sophisticated models may offer limited empirical benefits when applied to noisy and sparse clinical data.

- Sensitivity of models to the proportion of control samples: A crucial element in clinical research is the recruitment of subjects with appropriate study eligibility criteria. To better inform this process, models should undergo systematic evaluations regarding their sensitivity to different sample sizes and proportions of healthy samples.

- Dynamic updates of parameters: Our work shows substantial improvements in accuracy through iterative and dynamic parameter updating by incorporating evolving knowledge about the underlying data structure—an approach that can benefit other healthcare models where initial parameter estimation is uncertain.

- Importance of Bayesian priors given limited data: Incorporating and updating Bayesian priors, as shown in our work, is particularly important in healthcare settings with limited or imbalanced datasets.

2. Related Work

Event-based models (EBMs) were pioneered by Fonteijn et al. (2012) to uncover the progression sequence of neurodegenerative diseases from cross-sectional data. The fundamental premise of EBMs is that biomarkers become pathological in a consistent sequence across patients, with each disease stage reflecting the abnormality of an additional biomarker.

Several advancements have been made to the EBM framework. Some studies have focused on relaxing the strict ordinal ordering assumption. Temporal EBM (TEBM, Wijeratne et al., 2023), for instance, introduced a continuous representation of the ordering, while others have incorporated variability in ordering across subjects (Huang and Alexander, 2012; Venkatraghavan et al., 2019) or allowed for multiple central orderings (SuStaIn, Young et al., 2018). Further, a recently proposed model, the Parsimonious EBM (P-EBM, Cs et al., 2025) allows for some biomarkers to be affected simultaneously.

Additional enhancements have concentrated on the statistical underpinning of EBM. A critical limitation in early EBM implementations was the assumption of uniform distribution of disease stages. The original EBM by Fonteijn et al. (2012) and subsequent implementations by Young et al. (2014) and Firth et al. (2020) assumed equal probability for all disease stages—while in many observational cohorts such as ADNI later stages are typically underrepresented (Donohue et al., 2014).

Further constraints of the original EBM include assuming biomarker data follows Gaussian distributions and limited applicability to settings with a large number of biomarkers. KDE-EBM (Firth et al., 2020) employs Kernel Density Estimation to handle non-normal biomarker distributions, and Scaled EBM (sEBM; Tandon et al., 2023) and Variational EBM (vEBM; Wijeratne and Alexander, 2024) address challenges with high-dimensional biomarker data.

Despite these advancements, existing approaches such as EBM (Fonteijn et al., 2012), ALPACA (Huang and Alexander, 2012), DEBM (Venkatraghavan et al., 2019), SuStaIn (Young et al., 2018), KDE-EBM (Firth et al., 2020), TEBM (Wijeratne et al., 2023), sEBM (Tandon et al., 2023), vEBM (Wijeratne and Alexander, 2024) and P-EBM (Cs et al., 2025) typically estimate biomarker distributions once using methods like Gaussian Mixture Models (GMM) or Kernel Density Estimation (KDE), and then fix these parameters throughout inference, including during the Markov Chain Monte Carlo (MCMC) procedure. This static approach is hindered by a circular dependency in EBMs: accurate parameter estimation for biomarker distributions requires knowledge of the disease ordering and patient staging, but these are precisely what the algorithm aims to discover. Further, most existing approaches calculate data likelihood without modeling the distribution of disease stages, assuming equal representativeness of all disease stages.

Our work directly addresses these limitations by introducing a Stage-Aware EBM (SA-EBM) framework that dynamically updates biomarker distribution parameters and the disease stage distribution throughout MCMC. This approach leverages Bayesian principles to integrate evolving information about the underlying data structure, enabling more accurate biomarker ordering and patient staging across a range of scenarios.

3. Methods

3.1. Event-Based Modeling Framework

In the EBM framework, each biomarker exists in a "pre-event" or "post-event" state, with the "event" signifying the point at which the biomarker becomes pathological. Assuming a set of N biomarkers, we have N possible disease stages. Let J denote the total number of participants, j index participants, and k_j be their current disease stage, where $k_j = 0$ for healthy participants and $k_j > 0$ for progressing participants. Let n be a biomarker and $S(n)$ be its index (1-based) of the disease progression order $\boldsymbol{S}$. EBM assumes biomarker n becomes pathological when $k_j \geq S(n)$, with pre-event and post-event states modeled by separate distributions parameterized by $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ respectively.

Let $\boldsymbol{X}$ denote the full data, $\boldsymbol{X}_j$ be the biomarker measurements for participant j, and $x_{j,n}$ be biomarker n's measurement of participant j. The likelihood of $\boldsymbol{X}_j$ for a progressing participant with $k_j > 0$ is:

$$P(\boldsymbol{X}_j \mid \boldsymbol{S}, z_j = 1) = \sum_{k_j=1}^{N} P(k_j) p(\boldsymbol{X}_j \mid \boldsymbol{S}, z_j = 1, k_j) \tag{1}$$

where $P(k_j)$ is the prior probability of stage k_j, and z_j indicates this is a progression subject (otherwise $z_j = 0$). $p(\boldsymbol{X}_j \mid \boldsymbol{S}, k_j)$ is computed as:

$$p(\boldsymbol{X}_j \mid \boldsymbol{S}, z_j = 1, k_j) = \underbrace{\prod_{i=1}^{k_j} p(x_{j,S_i} \mid \boldsymbol{\theta}_{S_i})}_{\text{post-event likelihood}} \underbrace{\prod_{i=k_j+1}^{N} p(x_{j,S_i} \mid \boldsymbol{\phi}_{S_i})}_{\text{pre-event likelihood}} \tag{2}$$

where S_i is the i-th (1-based) biomarker to become pathological according to $\boldsymbol{S}$, and x_{j,S_i} is its measurement for participant j. The likelihood for a healthy participant is:

$$p(\boldsymbol{X}_j \mid \boldsymbol{S}, z_j = 0) = \prod_{i=1}^{N} p(x_{j,S_i} \mid \boldsymbol{\phi}_{S_i}) \tag{3}$$

The total likelihood of the dataset is:

$$P(\boldsymbol{X} \mid \boldsymbol{S}, \boldsymbol{z}) = \prod_{j=1}^{J} P(\boldsymbol{X}_j \mid \boldsymbol{S}, z_j) \tag{4}$$

where $\boldsymbol{z} = (z_1, z_2, ..., z_J)$. The goal of EBM is to find an $\boldsymbol{S}$ that maximizes the data likelihood. However, for a large N, exhaustive search of all possible orderings becomes computationally infeasible. In such cases, EBM employs Metropolis-Hastings MCMC to estimate the most probable ordering by proposing random swaps in the sequence and accepting or rejecting those proposals based on the resulting likelihood ratios.

3.2. Stage-Aware Event-Based Model (SA-EBM)

We introduce Stage-Aware EBM (SA-EBM) to address three major challenges in existing EBM implementations: First, the true parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ of the biomarker distributions

are unknown and must be estimated from the data. Second, the distribution of disease stages $(P(k_j))_{k_j=1}^N$ in the above equations is also unknown *a priori*. Third, the progression of the disease $\boldsymbol{S}$ is unknown and so, it will be estimated from the data simultaneously.

Unlike previous EBMs that use static $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ estimated without taking into account the likely disease stages of participants, or update the disease stage distribution during MCMC without a proper Bayesian prior, SA-EBM considers $(P(k_j))_{k_j=1}^N$ as drawn from a Dirichlet distribution, and iteratively update $\boldsymbol{\phi}$, $\boldsymbol{\theta}$, and $(P(k_j))_{k_j=1}^N$ based on the evolving best estimates from the sampler. Detailed procedures are displayed in Algorithm 1.

Algorithm 1 Stage-Aware Event-Based Model (SA-EBM) Algorithm

1: $\boldsymbol{\pi} = (P(k_j))_{k_j=1}^N \sim \text{Dirichlet}(\boldsymbol{\alpha}_0)$, where $\boldsymbol{\alpha}_0 = \mathbf{1}_N$
2: $\boldsymbol{\theta} = (\boldsymbol{\theta}_n)_{n=1}^N$ (post-event state) and $\boldsymbol{\phi} = (\boldsymbol{\phi}_n)_{n=1}^N$ (pre-event state) using K-Means clustering and conjugate prior updates on the biomarker data
3: Initialize $\boldsymbol{S}$ as sampled uniformly from all permutations: $\boldsymbol{S} \sim \text{Uniform}(N!)$.
4: $\ell = -\infty$
5: **for** $i = 1$ to M (number of MCMC iterations) **do**
6: Propose $\boldsymbol{S}'$ by randomly swapping two biomarkers in $\boldsymbol{S}$.
7: $\mathbf{A} = (P(k_j \mid \boldsymbol{X}_j, \boldsymbol{S}', \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}) \quad \forall k_j \in \{1, 2, ..., N\})_{j=1}^J$.
8: Compute $\boldsymbol{\theta}', \boldsymbol{\phi}'$ based on $\boldsymbol{S}'$ and $\mathbf{A}$
9: $\ell' = \mathcal{L}(\boldsymbol{X} \mid \boldsymbol{S}', \boldsymbol{\theta}', \boldsymbol{\phi}', \boldsymbol{\pi})$ using Equation 4
10: $p = \min(1, \exp(\ell' - \ell))$
11: $U \sim \text{Uniform}(0, 1)$
12: **if** $U < p$ **then**
13: $\boldsymbol{S} \leftarrow \boldsymbol{S}'$
14: $\ell \leftarrow \ell'$
15: $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}'$ and $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi}'$
16: $\mathbf{A} \leftarrow (P(k_j \mid \boldsymbol{X}_j, \boldsymbol{S}', \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}) : k_j \in \{1, 2, ..., N\})_{j=1}^J$
17: $\boldsymbol{\pi} \leftarrow \boldsymbol{\pi}' \sim \text{Dirichlet}\left([\boldsymbol{\alpha}_{0k_j} + \sum_{j=1}^J A_{j,k_j}]_{k_j=1}^N\right)$
18: **end if**
19: **end for**
20: **Return** $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}, \mathbf{A}, (\ell_m)_{m=1}^M$, and $(\boldsymbol{S}_m)_{m=1}^M$

Note: When the variant is Hard K-Means, lines 7, 8, and 15 do not apply, and line 9 becomes $\ell' = \mathcal{L}(\boldsymbol{X} \mid \boldsymbol{S}', \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi})$

Based on Bayes' rule, for all $k_j \in \{1, 2, ..., N\}$, we have

$$P(k_j \mid \boldsymbol{X}_j, \boldsymbol{S}', \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}) \propto P(k_j \mid \boldsymbol{\pi}) P(\boldsymbol{X}_j \mid k_j, \boldsymbol{S}', \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}) \tag{5}$$

where $P(k_j \mid \boldsymbol{\pi}) = \pi_{k_j}$ and $P(\boldsymbol{X}_j \mid k_j, \boldsymbol{S}', \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi})$ can be calculated using Eq. 1.

We implement five different approaches to estimating and updating biomarker distribution parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ within the SA-EBM framework. Details are available in Appendix B.

4. Synthetic Experiments

We designed a series of controlled synthetic experiments to evaluate the performance of SA-EBM against existing event-based modeling (EBM) algorithms. Our goal was to assess both the accuracy of inferred biomarker orderings and subjects' disease stages under a wide range of realistic conditions, including ordinal vs. continuous disease stages following uniform vs. non-uniform distributions, biomarker data following normal vs. non-normal biomarker distributions, and varying participant sizes and healthy group percentages (ratios). We generated synthetic data using two distinct models: an EBM-native model based on Fonteijn et al. (2012) and a sigmoid model adapted from Venkatraghavan et al. (2019).

4.1. EBM-Native Generative Model with Ordinal k_j

The EBM-native model simulates the core assumptions of the EBM framework. Measurement of biomarker n in subject j, i.e., $x_{n,j}$, is generated as follows:

1. Stage Assignment: A disease stage k_j is assigned to each subject. Healthy subjects are assigned $k_j = 0$. For impacted subjects:
   - A stage distribution $\boldsymbol{\pi}$ is sampled from a Dirichlet distribution: $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\alpha})$. This distribution represents the probability of an impacted participant being in each disease stage.
   - The number of subjects in different disease stages, represented by $\boldsymbol{k}$, is drawn from a Multinomial distribution: $\mathbf{k} \sim \text{Multinomial}(\text{J}_{\text{impacted}}, \boldsymbol{\pi})$, where J_{impacted} is the number of impacted subjects; therefore, $\sum_{i=1}^{N} k_i = J_{\text{impacted}}$.
   - Generate a sequence of J_{impacted} disease stages based on $\boldsymbol{k}$ and concatenate it with J_{healthy} instances of $k_j = 0$. The combined sequence of $J = J_{\text{impacted}} + J_{\text{healthy}}$ is uniformly randomized, determining the final stage $k_j \in \{0, 1, \ldots, N\}$ for each participant.
2. Biomarker Generation: For each biomarker n, if $S(n) \leq k_j$, $x_{n,j}$ is generated from the post-event distribution with parameters $\boldsymbol{\theta}_n = (\mu_{n,\theta}, \sigma^2_{n,\theta})$; Otherwise, the pre-event distribution with parameters $\boldsymbol{\phi}_n = (\mu_{n,\phi}, \sigma^2_{n,\phi})$.

Mathematically:

$$
\begin{aligned}
x_{n,j} \mid \boldsymbol{S}, k_j, \boldsymbol{\theta}_n, \boldsymbol{\phi}_n, z_j \sim I(z_j = 1) \Big[I(S(n) \leq k_j)\, p(x_{n,j} \mid \boldsymbol{\theta}_n) + I(S(n) > k_j)\, p(x_{n,j} \mid \boldsymbol{\phi}_n) \Big] \\
+ (1 - I(z_j = 1))\, p(x_{n,j} \mid \boldsymbol{\phi}_n)
\end{aligned}
\tag{6}
$$

where $\boldsymbol{S} \sim \text{Uniform}(N!)$ is a discrete variable following a distribution of uniform permutation. This permutation is randomized for each dataset. The graphical model of this generative process is presented in Figure 1.

For stage distribution $\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\alpha})$, we tested two configurations:

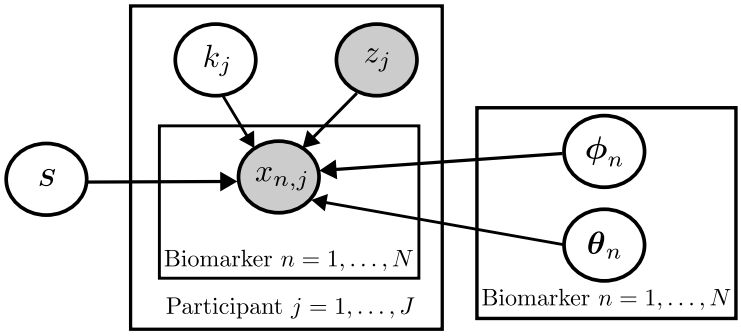

Figure 1: **Graphical model of EBM**

1. $\boldsymbol{\alpha} = 100_N$, approximating a uniform distribution.
2. A specific $\boldsymbol{\alpha}$ to mimic a normal distribution.

For biomarker measurements, we used both normal and non-normal distributions (with details available in Table 3 and Figure 4 in Appendix C and E):

- Normal Distributions: Parameters were estimated from ten biomarkers related to Alzheimer's disease reported in Chen et al. (2016).
- Non-Normal Distribution: Custom mixture distributions were designed to capture irregular, non-Gaussian behaviors.

We conducted four experiments with data generated from the EBM-native model:
Experiment 1: $\boldsymbol{S}$ & Ordinal k_j (Dirichlet-Multinomial with $\boldsymbol{\alpha}$ mimicking a normal distribution) & Normal $x_{n,j}$ with fixed parameters.
Experiment 2: $\boldsymbol{S}$ & Ordinal k_j (Dirichlet-Multinomial with $\boldsymbol{\alpha}$ mimicking a normal distribution) & Non-Normal $x_{n,j}$.
Experiment 3: $\boldsymbol{S}$ & Ordinal k_j (Dirichlet-Multinomial with $\alpha_i = 100 \quad \forall i$, mimicking a uniform distribution) & Normal $x_{n,j}$ with fixed parameters.
Experiment 4: $\boldsymbol{S}$ & Ordinal k_j (Dirichlet-Multinomial with $\alpha_i = 100 \quad \forall i$, mimicking a uniform distribution) & Non-Normal $x_{n,j}$.

4.2. Sigmoid Model with Continuous k_j

We also used a modified version of the generative model from Venkatraghavan et al. (2019), based on the simulation framework by Young et al. (2015). This model assumes that biomarker values for healthy individuals follow normal distributions, while those for progressing individuals deviate monotonically from healthy values over time. In this model, k_j is continuous. The differences between our model and that used in Venkatraghavan et al. (2019) are: First, we introduce directional variability by randomly flipping the sign of the sigmoid trajectory per biomarker; Second, we assume a global progression order, ensuring that all individuals share the same event times within a given experiment. Biomarker values are generated as follows:
When $k_j = 0$,

$$x_{n,j} \sim \mathcal{N}(\mu_{n,\phi}, \sigma^2_{n,\phi})$$

When $k_j > 0$,

$$x_{n,j} \sim \mathcal{N}(\mu_{n,\phi}, \sigma^2_{n,\phi}) + \frac{(-1)^{I_n} R_n}{1 + e^{-\rho_n(k_j - \xi_n)}}$$

$R_n = \mu_{n,\theta} - \mu_{n,\phi}$ is the range of a biomarker. $\rho_n = \max\left(1, \frac{|R_n|}{\sqrt{\sigma^2_{n,\theta} + \sigma^2_{n,\phi}}}\right)$ controls the slope. $I_n \sim$ Bernoulli(0.5) randomly flips the direction of progression of the biomarker n. The ideal sigmoid transitions for all biomarkers are analyzed and visualized in Appendix G.

We explored both ordinal ($\boldsymbol{S}$) and continuous ($\boldsymbol{\xi}$) formulations of event time. Let $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)$ denote the vector of event times, where ξ_n is the event time associated with biomarker n, drawn from a scaled Beta distribution with parameters described below:

Experiment 5: $\boldsymbol{S}$ & Continuous k_j (scaled Beta distribution, $\lambda = N, \alpha = \beta = 1$, approximating uniform).

Experiment 6: $\boldsymbol{S}$ & Continuous k_j (scaled Beta distribution, $\lambda = N, \alpha = 5, \beta = 2$).

Experiment 7: $\boldsymbol{\xi}$ (Scaled Beta distribution, $\lambda = N, \alpha = \beta = 2$, approximating normal) & Continuous k_j (scaled Beta distribution, $\lambda = N, \alpha = \beta = 1$, approximating uniform).

Experiment 8: $\boldsymbol{\xi}$ (Scaled Beta distribution, $\lambda = N, \alpha = \beta = 2$, approximating normal) & Continuous k_j (scaled Beta distribution, $\lambda = N, \alpha = 5, \beta = 2$).

We added variability to $\xi_{n,j}$ to account for individual differences in event times: $\xi_{n,j} = \text{clip}\,(\xi_n + \delta,\ 0,\ N)$, where $\delta \sim \mathcal{N}(0, N \cdot 0.05)$. $N = 10$ in our experiments ensured 95% of the noise fell within $[-1, +1]$. This experiment is designed to be closer to real-world datasets and to provide a fair comparison with DEBM (Venkatraghavan et al., 2019).

Experiment 9: $\boldsymbol{\xi}$ (Scaled Beta distribution, $\lambda = N, \alpha = \beta = 2$, approximating normal) with added noise & Continuous k_j (scaled Beta distribution, $\lambda = N, \alpha = 5, \beta = 2$).

In Appendix E, Figure 4 visualizes the pre- and post-event distributions for each biomarker in theoretical normal distributions, non-normal distributions, and the sigmoid model (a dataset of Experiment 9). Table 4 provides a summary of configurations of all experiments.

4.3. Experiment Setup

For each experiment, we varied the total numbers of participants ($J = 50, 200, 500, 1000$) and healthy ratios ($r = 0.1, 0.25, 0.5, 0.75, 0.9$), creating 50 random datasets per configuration. Each dataset includes both healthy and progressing participants, with known ground truth for $\boldsymbol{S}$ and k_j. In total, we generated 9,000 datasets (9 experiments $\times$ 4 participant sizes $\times$ 5 healthy ratios $\times$ 50 repetitions).

We evaluated our five SA-EBM variants (Hard K-Means, Conjugate Priors, MLE, EM, and KDE) against established algorithms selected to represent the state-of-the-art in event-based modeling:

1. EBM with GMM: We included two independent implementations of EBM approach using Gaussian Mixture Models:

- UCL GMM: The implementation from the UCL POND research group (Firth et al., 2020)
- DEBM GMM: The GMM-based implementation released alongside DEBM (Venkatraghavan et al., 2019)

2. DEBM: The discriminative Event-Based Model by Venkatraghavan et al. (2019), which was specifically designed to handle subject-specific variations in event order and incorporates updates to staging probabilities.

3. KDE-EBM (UCL KDE): The nonparametric KDE implementation by Firth et al. (2020) designed to handle non-Gaussian biomarker distributions.

These benchmark algorithms were selected to represent diverse approaches within the EBM framework, covering different parameter estimation techniques (parametric vs. non-parametric), different assumptions about event ordering (fixed vs. variable), and different approaches to staging probability estimation (static vs. dynamic). Meanwhile, all these methods are fixed-parameter approaches in the sense that distribution parameters are estimated once (e.g., via GMM or KDE) and kept fixed throughout inference. This design choice makes them ideal baselines to contrast with our stage-aware model.

All algorithms were run with 10,000 MCMC iterations, except for DEBM which employs a different inference algorithm. For methods requiring an initialization phase for $\boldsymbol{\phi}, \boldsymbol{\theta}$ estimation (DEBM GMM, UCL GMM, and UCL KDE), we used 10 initializations with 1,000 EM iterations each, which is more than demonstrated by POND (2025).

4.4. Evaluation Metrics

We evaluated algorithm performance using two main metrics: (1) the accuracy of biomarker ordering and (2) the accuracy of patient staging.

For ordering accuracy, we employed normalized Kendall's Tau distance: a standard metric in progression modeling (Young et al., 2023; Tandon et al., 2023; Cs et al., 2025) that measures the distance between two ordered sequences. Normalized Kendall's Tau distance ranges from 0 (perfect match) to +1 (inverse order). We measured the distance between the ordering picked by the model and the real ordering. SA-EBM selected the biomarker ordering that maximized the data log-likelihood. Benchmark algorithms were evaluated based on the orderings they directly produced.

For staging accuracy, we used mean absolute error (MAE) to quantify the average deviation between predicted and true participants' stages. For experiments with continuous ground truth stages, we converted these to ordinal positions by finding the appropriate insertion point within the sorted sequence of event times. For our SA-EBM algorithms, after MCMC iterations, we had obtained the ordering with the highest data log-likelihood $\boldsymbol{S}_{\max}$, and final $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}$. Based on these, we calculated the staging posterior: $(P(k_j \mid \boldsymbol{X}_j, \boldsymbol{S}_{\max}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}) \quad \forall k_j \in \{0, 1, 2, ..., N\})_{j=1}^{J}$. Note that we ignored the ground truth of diagnosis labels, i.e., healthy or impacted here. We then sampled k_j from a discrete distribution $P(k_j)$ using weighted random selection: $k_j \sim \text{Categorical}(P(k_j))$.

5. Synthetic Experiments Results

We conducted all experiments on the CHTC cluster at the University of Wisconsin-Madison (Center for High Throughput Computing, 2006), completing them in approximately 18 hours. In Experiment 2, the UCL KDE implementation failed on 21 datasets due to singular matrix errors, which are not handled in the algorithm by POND (2025).

5.1. Overall Performance

As shown in Figures 2 and 5, SA-EBM algorithms produced higher accuracy scores compared to the benchmark methods on both ordering and staging tasks. As shown in Figure 7 and 8 in Appendix F, among the five SA-EBM variants, the Conjugate Priors approach achieved the highest average ordering accuracy with a normalized Kendall's Tau distance of 0.18 ± 0.01 (95% CI), followed by MLE (0.18 ± 0.01) and EM (0.19 ± 0.01). The KDE (0.24 ± 0.01) and Hard K-Means (0.25 ± 0.01) variants showed moderate performance. The benchmark algorithms (UCL KDE, DEBM GMM, DEBM, and UCL GMM) displayed lower performance with average normalized τ distances above 0.32.

Only Conjugate Priors (0.91 ± 0.03) and MLE (0.92 ± 0.03) achieved average MAE values below 1.00. Results of DEBM GMM (1.22 ± 0.05), Hard K-Means (1.22 ± 0.05), DEBM (1.24 ± 0.05), EM (1.29 ± 0.11) and KDE (1.37 ± 0.08) were below 1.50. UCL GMM (1.56 ± 0.05) and UCL KDE (1.80 ± 0.09) had the worst average MAE values.

5.2. Results on Performance Across Experimental Conditions

Figures 2 and 5 (Appendix F) present detailed performance breakdowns across experimental configurations, sample sizes (J) and healthy ratios (r).

Sample Size: Ordering performance generally improved with an increasing sample size across all algorithms and all experiments. The most substantial improvements for SA-EBM occurred between $J = 50$ and $J = 200$, with smaller incremental gains observed at $J = 500$ and $J = 1,000$. Sample size does not influence staging performance very much.

Proportion of Control Samples: As the healthy ratio (r) increased, the ordering performance of benchmark algorithms decreased substantially. In contrast, SA-EBM maintained relatively stable performances across all proportions of control samples. At $r = 0.9$ and $J = 500$ (corresponding to only 50 progressing subjects), our method achieved performance comparable to settings with more impacted participants (e.g., $r = 0.1, J = 200$). As for the performance on staging tasks, the accuracy of SA-EBM improved as the healthy ratio increased. Accuracy trends of benchmark algorithms varied by the specific algorithm. For example, staging accuracy of DEBM GMM and DEBM in general improved as the healthy ratio increased but showed a downward trend in some experiments. Staging accuracy of UCL GMM decreased with higher healthy ratios whereas UCL KDE showed an "inverted V" curve.

Biomarker Distribution: As shown in Figure 7 and 8 in Appendix F, in Experiment 2 and 4, the parametric variants of our SA-EBM approach (Conjugate Priors, MLE, EM) consistently outperformed the nonparametric KDE variants even with non-Gaussian data.

Progression Model: When tested on data from continuous progression models (Experiments 5-9), our SA-EBM variants maintained performance levels similar to those observed

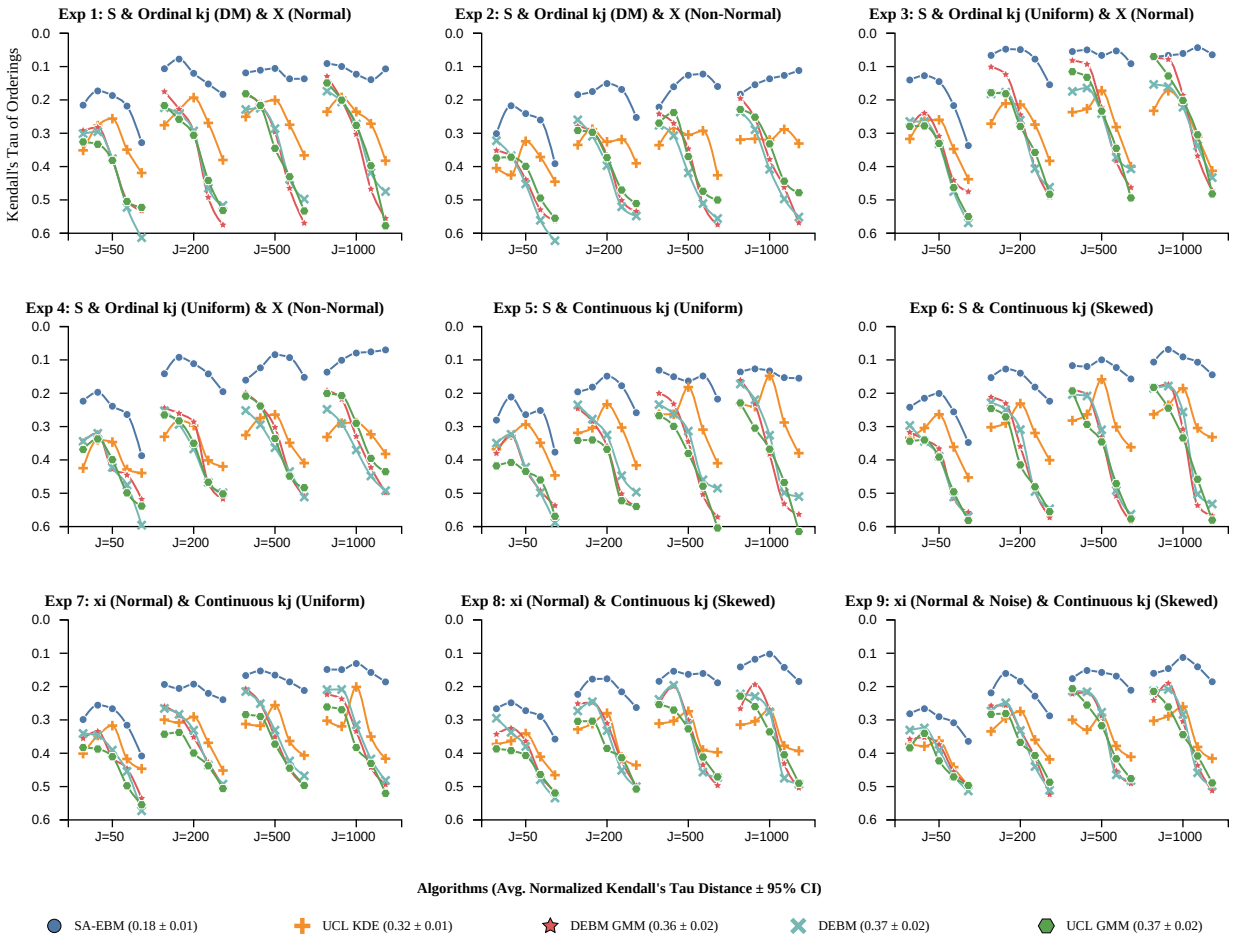

Figure 2: **Average normalized Kendall's Tau distance values (±95% CI)**. We use Conjugate Priors to represent SA-EBM as it has the best performance. Each panel represents a different experimental configuration with varying data generation models, stage distributions, and biomarker distributions. The x-axis within each panel shows different participant sizes ($J = 50, 200, 500, 1000$). Within each participant size are different healthy ratios (r), i.e., the percentage of healthy participants among all subjects. From left to right are $r = 0.1, 0.25, 0.5, 0.75, 0.9$. The y-axis shows the normalized Kendall's Tau distance (lower is better). Data points represent mean performance across 50 variants of the same experimental configuration, sample size, and healthy ratio. SA-EBM (Conjugate Priors) consistently outperform static and baseline methods. Performance generally improves with increasing sample size, while fixed-parameter methods degrade under high healthy ratios and non-Gaussian data.

with the original EBM model for data similar to its model assumptions (Experiments 1-4). This was consistent across both the continuous uniform stage distributions (Experiment 5 & 7) and the continuous skewed distributions (Experiment 6, 8, & 9).

Individual Variability: In Experiment 9, which incorporated subject-level variability in event times, our SA-EBM variants showed only a minor decrease in performance compared to Experiment 8, which has the same configurations except for the perturbations to event times.

5.3. Algorithm-Specific Performance Patterns

Among the five SA-EBM variants, Conjugate Priors showed the best performance. The Hard K-Means approach, which represents a static parameter estimation strategy similar to existing methods but with our staging probability updates, outperformed almost all benchmark algorithms in both ordering and staging tasks but lagged behind our dynamic parameter updating variants. The benchmark KDE-EBM implementation (UCL KDE) consistently underperformed our KDE variant across all conditions and both tasks.

Figure 6 (Appendix F) displays the average performances and variability thereof for all tested algorithms by aggregating their results across all 9,000 datasets. It clearly shows that existing EBM implementations have lower ordering accuracies and higher variability. It should be noted that DEBM and DEBM GMM performed well on the staging task.

6. Real World Dataset

We applied our SA-EBM algorithms to real-world data from the Alzheimer's Disease Neuroimaging Initiative (ADNI, Mueller et al., 2005). 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).

We documented how we processed the ADNI data in Appendix H. The final dataset included 726 participants, distributed across diagnostic categories as follows: AD (153, 21.1%), LMCI (236, 32.5%), CN (155, 21.3%), and EMCI (182, 25.1%). Among them, 413 (56.9%) were men 313 (43.1%) were women. Age distribution can be found in Figure 11 (Appendix I). These participants came from the following study protocols: ADNI1 (275, 37.9%), ADNIGO (76, 10.5%), and ADNI2 (375, 51.7%).

Since Conjugate Priors and MLE had the best performances based on results of the synthetic experiments, we applied these two algorithms to ADNI three times. We picked the result with the largest data log-likelihood, which was from Conjugate Priors. We also applied UCL GMM, DEBM, and DEBM GMM to ADNI for comparison as the number of participants enrolled in ADNI studies has increased since these methods were published. We failed to run UCL KDE on ADNI due to "singular matrix" error.

The result from Conjugate Priors (Figure 3) suggests that ventricular enlargement occurs first, followed by cognitive decline (RAVLT Immediate, ADAS, and MMSE). Next, pathology in $A\beta_{1-42}$ protein and the two Tau-related biomarkers. Neurodegeneration in brain

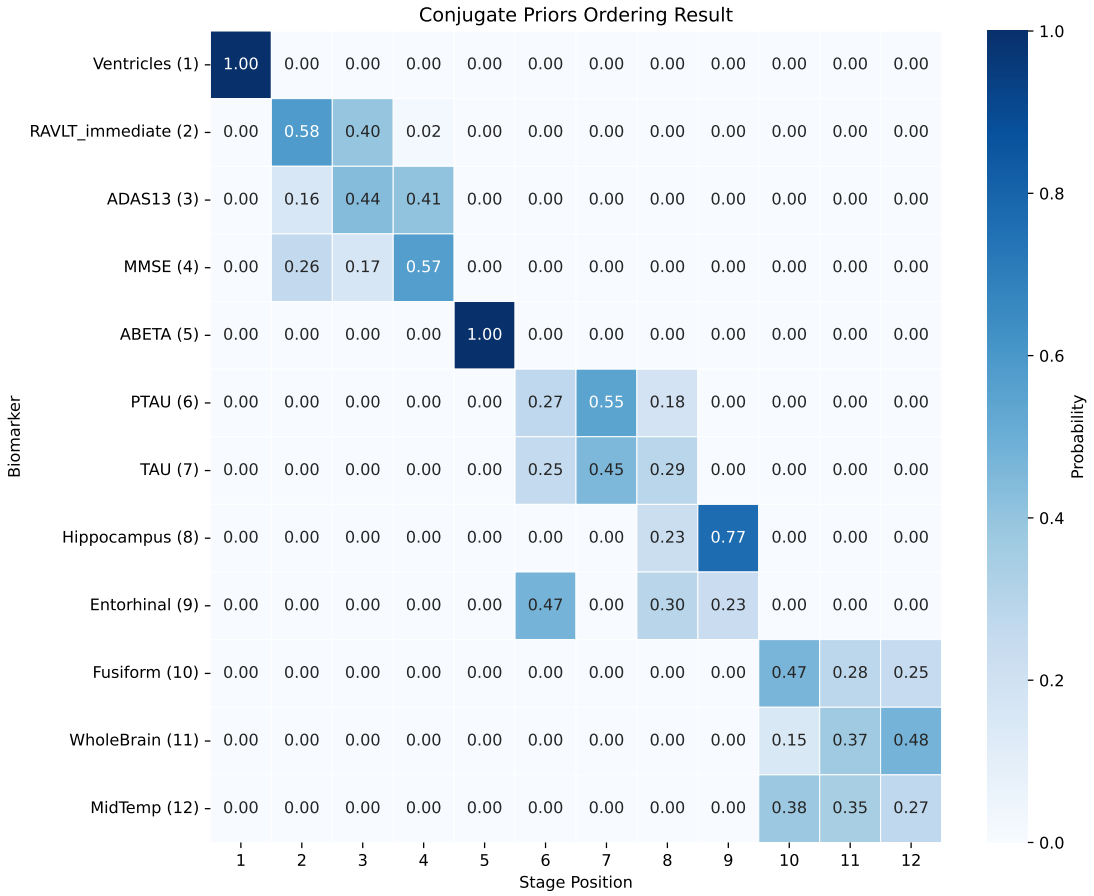

Figure 3: **Ordering Result with Conjugate Priors on ADNI data**: The heatmap shows uncertainties for 10,000 MCMC iterations with 500 burn-in and no thinning. The number inside the parenthesis in the Y-axis indicates the result according to the ordering associated with the largest log data likelihood. Each cell indicates the probability of each biomarker getting affected in a specific stage, according to the results from the last 9,500 MCMC iterations.

regions—including the Hippocampus, Entorhinal cortex, Fusiform gyrus, WholeBrain, and MidTemporal area—occurs last.

In contrast, as shown in Appendix K, UCL GMM (Figure 13) identifies this sequence: brain volume loss, amyloid pathology, ventricular enlargement, cognitive decline, tau pathology, and again brain volume loss. DEBM GMM (Figure 14) produces the following progression: abnormalities in $A\beta_{1-42}$ and tau proteins, followed by ventricular enlargement, cognitive decline, and then brain atrophy. DEBM (Figure 15) identifies another sequence: amyloid pathology, cognitive decline, tau pathology, further cognitive declines, another tau biomarker, and lastly brain atrophy and ventricular enlargement.

7. Discussion

Our results clearly demonstrate the benefits of the Stage-Aware EBM (SA-EBM). By dynamically updating distributions of disease stages and biomarker measurements, our algorithms exhibit improved robustness across a wide range of challenging scenarios compared to the prior EBM algorithms.

The advantages that SA-EBM has on both ordering and staging tasks are most evident when the proportion of healthy participants (r) is high. When the number of participants (J) is large and the ratio of healthy participants (r) is small, SA-EBM approaches still tend to outperform other methods, but the performance advantage is smaller. This is likely due to the computational problem being easier when there are more reliable data from impacted participants. However, when J is small and r is large, benchmark algorithms show substantially reduced performance. This robustness of our algorithms has important implications for clinical study design, particularly for rare diseases where recruiting large numbers of impacted participants is challenging.

The SA-EBM ordering results indicate that while larger sample sizes generally improve performance, the benefits may saturate when sample sizes are in the range of 200 to 1000 participants. This may have practical implications for clinicians and researchers. It suggests that researchers may perform reliable progression modeling even when sample sizes are limited. However, we caution against over-interpretation of this finding as real-world clinical studies are more complex than the synthetic benchmarks we conducted.

Interestingly, KDE-based algorithms showed reduced performance compared to those relying on Gaussian assumptions on experiments that used non-Gaussian biomarker distributions. This seemingly counterintuitive result can be attributed to the bias-variance tradeoff. With limited data, KDE algorithms tend to overfit, while Gaussian assumption-based algorithms act as a form of regularization. This explains why Conjugate Priors, MLE and EM outperform KDE and UCL KDE. For clinical applications, this suggests that parametric approaches may be preferable in many practical scenarios, despite their theoretical limitations with non-Gaussian data.

Furthermore, our algorithms exhibited robustness to data deviating from its assumptions. In the experiments with continuous event times and stages, and even with subject-specific variations in event times, SA-EBM methods, especially the variants of Conjugate Prior and MLE, still outperformed other methods. While the EBM is designed for ordinal events and staging, its ability to accurately estimate pre- and post-event distributions, as well as stage distributions, allows it to perform well even with data generated from a continuous sigmoid model. This suggests that SA-EBM may provide accurate results regardless of whether the disease progresses in discrete stages or along a continuous trajectory.

Our result on the ADNI dataset is different from that presented in Young et al. (2014): tau and amyloid pathology first, followed by brain atrophy, and then cognitive impairment, and lastly volumetric measures of brain regions. It is also different from that reported in Archetti et al. (2019) using DEBM: $A\beta_{1-42}$ protein, cognitive scores, tau pathology, and brain region atrophy. These disparities might be due to the differences in the dataset. For example, Young et al. (2014), published more than ten years ago with limited ADNI data, had only 285 participants. Also, Archetti et al. (2019) included participants with missing data. Additionally, whereas both Young et al. (2014) and Archetti et al. (2019) performed

log transformations on Tau and P-Tau measurements to improve data normality, we used the original data. As shown in the Figure 16 (Appendix K), our method provides similar results when these quantities are log transformed.

We interpret our result in Figure 3 as follows. First, biomarkers of the same type are grouped nicely with uncertainty of the order within each type. For example, the three biomarkers representing cognitive scores are grouped together and the uncertainties show they become pathological roughly simultaneously. The same happens to amyloid and tau proteins, and brain volumes. Second, our Ventricles → Cognition (C) → Amyloid (A) → Tau (T) → Neurodegeneration (N) ordering is partially consistent with the ATNC ordering that is the basis for the revised criteria of Alzheimer's Association Workgroup 2024 (AA-2024; Jack Jr et al., 2024). The discrepancy between our inferred ordering and the canonical progression may reflect the heterogeneity of Alzheimer's disease, as not all individuals adhere strictly to the sequence outlined in the ATNC framework Mendes et al. (2025).

The staging result (Figure 17, Appendix L) and the trace plot (Figure 12, Appendix J) provide additional support for the ordering estimated by our SA-EBM algorithm. Specifically: (1) Control and EMCI participants were predominantly assigned to the first three stages; (2) AD participants were mostly assigned to the later stages; and (3) The log-likelihood increased and eventually converged. In contrast, the staging results of the benchmark algorithms appeared problematic. As shown in the Appendix L, UCL GMM (Figure 18) assigned CN participants to late stages. DEBM (Figure 20) and DEBM GMM (Figure 19) performed well, but they assigned an excessive number of non-CN participants to stage 0 or assigned AD patients to early stages. Nevertheless, we caution that staging accuracy does not necessarily imply correctness of the estimated ordering. For instance, in our synthetic experiments, both DEBM and DEBM GMM achieve strong staging performance but fall short in accurately recovering the true event ordering.

7.1. Limitations

While our results are promising, several limitations are acknowledged. Numerically simulated datasets, which—though carefully designed to mimic real-world scenarios—may still not capture the full complexity of clinical data, including confounders, missing data, and measurement errors.

SA-EBM assumes a single global biomarker progression sequence shared across all participants and is primarily concerned with ordinal order without modeling actual temporal intervals between events. This fails to capture the complexity of real-world disease progression. Besides, we recognize that a lack of variations in event times at the subject level might have influenced the performance of DEBM, which is based on the assumption that each subject may have a different ordering. To make fair comparisons, we have added the Experiment 9 where perturbations to event times are applied. The results remain consistent with prior experiments.

Lastly, compared to fixed-parameter methods, our approach incurs a higher computational overhead due to iterative updates. That said, we do not believe this poses a serious bottleneck. A comprehensive runtime analysis, presented in Appendix M, demonstrates that while SA-EBM is indeed slower than benchmark algorithms, its runtime remains acceptable for the majority of medical and research contexts.

Acknowledgments

We thank the CHTC at the University of Wisconsin-Madison for computing support. JLA was funded by the Japan Probabilistic Computing Consortium Association. VP, VAN, NA, and JLA were supported in part by NIH grants R01NS123378, R01NS111022, R01NS105646, R01NS117568, and P50HD105353.

Acknowledgments to ADNI can be found in Appendix O.

References

Damiano Archetti, Silvia Ingala, Vikram Venkatraghavan, Viktor Wottschel, Alexandra L Young, Maura Bellio, Esther E Bron, Stefan Klein, Frederik Barkhof, Daniel C Alexander, et al. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in alzheimer's disease. *NeuroImage: Clinical*, 24:101954, 2019.

Center for High Throughput Computing. Center for high throughput computing, 2006. URL https://chtc.cs.wisc.edu/.

Guangyu Chen, Hao Shu, Gang Chen, B Douglas Ward, Piero G Antuono, Zhijun Zhang, Shi-Jiang Li, Alzheimer's Disease Neuroimaging Initiative, et al. Staging Alzheimer's disease risk by sequencing brain function and structure, cerebrospinal fluid, and cognition biomarkers. *Journal of Alzheimer's Disease*, 54(3):983–993, 2016.

Parker Cs, NP Oxtoby, DC Alexander, H Zhang, AL Young the Alzheimer's Disease Neuroimaging Initiative, et al. Parsimonious EBM: generalising the event-based model of disease progression for simultaneous events. *NeuroImage*, page 121162, 2025.

Michael C Donohue, Hélène Jacqmin-Gadda, Mélanie Le Goff, Ronald G Thomas, Rema Raman, Anthony C Gamst, Laurel A Beckett, Clifford R Jack Jr, Michael W Weiner, Jean-François Dartigues, et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia*, 10:S400–S410, 2014.

Arman Eshaghi, Razvan V Marinescu, Alexandra L Young, Nicholas C Firth, Ferran Prados, M Jorge Cardoso, Carmen Tur, Floriana De Angelis, Niamh Cawley, Wallace J Brownlee, et al. Progression of regional grey matter atrophy in multiple sclerosis. *Brain*, 141(6): 1665–1677, 2018.

Nicholas C Firth, Silvia Primativo, Emilie Brotherhood, Alexandra L Young, Keir XX Yong, Sebastian J Crutch, Daniel C Alexander, and Neil P Oxtoby. Sequences of cognitive decline in typical Alzheimer's disease and posterior cortical atrophy estimated using a novel event-based model of disease progression. *Alzheimer's & Dementia*, 16(7):965–973, 2020.

Hubert M Fonteijn, Marc Modat, Matthew J Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z Hobbs, Rachael I Scahill, Sarah J Tabrizi, Sebastien Ourselin, Nick C Fox, et al. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage*, 60(3):1880–1889, 2012.

Andrew Gelman, Daniel Simpson, and Michael Betancourt. The prior can often only be understood in the context of the likelihood. *Entropy*, 19(10):555, 2017.

Bo Hu, Ying Yu, Xin-Wen Yu, Min-Hua Ni, Yan-Yan Cui, Xin-Yu Cao, Ai-Li Yang, Yu-Xin Jin, Sheng-Ru Liang, Si-Ning Li, et al. Sequence of episodic memory-related behavioral and brain-imaging abnormalities in type 2 diabetes. *Nutrition & Diabetes*, 15(1):1, 2025.

Jonathan Huang and Daniel Alexander. Probabilistic event cascades for Alzheimer's disease. *Advances in neural information processing systems*, 25, 2012.

Clifford R Jack Jr, J Scott Andrews, Thomas G Beach, Teresa Buracchio, Billy Dunn, Ana Graf, Oskar Hansson, Carole Ho, William Jagust, Eric McDade, et al. Revised criteria for diagnosis and staging of alzheimer's disease: Alzheimer's association workgroup. *Alzheimer's & Dementia*, 20(8):5143–5169, 2024.

Augusto J Mendes, Federica Ribaldi, Michela Pievani, Cecilia Boccalini, Valentina Garibotto, Giovanni B Frisoni, and Alzheimer's Disease Neuroimaging Initiative. Validating the amyloid cascade through the revised criteria of alzheimer's association workgroup 2024 for alzheimer disease. *Neurology*, 104(11):e213675, 2025.

Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics*, 15(4):869–877, 2005.

UCL POND. Kernel Density Estimation Event-Based Model (kde_ebm). https://github.com/ucl-pond/kde_ebm, 2025. Software repository. Last updated February 2025.

Raghav Tandon, Anna Kirkpatrick, and Cassie S Mitchell. sEBM: Scaling event based models to predict disease progression via implicit biomarker selection and clustering. In *International Conference on Information Processing in Medical Imaging*, pages 208–221. Springer, 2023.

Vikram Venkatraghavan, Esther E Bron, Wiro J Niessen, Stefan Klein, Alzheimer's Disease Neuroimaging Initiative, et al. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. *NeuroImage*, 186:518–532, 2019.

Peter Wijeratne and Daniel Alexander. Unscrambling disease progression at scale: fast inference of event permutations with optimal transport. *Advances in Neural Information Processing Systems*, 37:63316–63341, 2024.

Peter A Wijeratne, Arman Eshaghi, William J Scotton, Maitrei Kohli, Leon Aksman, Neil P Oxtoby, Dorian Pustina, John H Warner, Jane S Paulsen, Rachael I Scahill, et al. The temporal event-based model: Learning event timelines in progressive diseases. *Imaging Neuroscience*, 1:1–19, 2023.

Alexandra L Young, Neil P Oxtoby, Pankaj Daga, David M Cash, Nick C Fox, Sebastien Ourselin, Jonathan M Schott, and Daniel C Alexander. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain*, 137(9):2564–2577, 2014.

Alexandra L Young, Neil P Oxtoby, Sebastien Ourselin, Jonathan M Schott, Daniel C Alexander, Alzheimer's Disease Neuroimaging Initiative, et al. A simulation system for biomarker evolution in neurodegenerative disease. *Medical image analysis*, 26(1):47–56, 2015.

Alexandra L Young, Razvan V Marinescu, Neil P Oxtoby, Martina Bocchetta, Keir Yong, Nicholas C Firth, David M Cash, David L Thomas, Katrina M Dick, Jorge Cardoso, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nature communications*, 9(1):4273, 2018.

Alexandra L Young, Leon M Aksman, Daniel C Alexander, Peter A Wijeratne, and Alzheimer's Disease Neuroimaging Initiative. Subtype and stage inference with timescales. In *International Conference on Information Processing in Medical Imaging*, pages 15–26. Springer, 2023.

Alexandra L Young, Neil P Oxtoby, Sara Garbarino, Nick C Fox, Frederik Barkhof, Jonathan M Schott, and Daniel C Alexander. Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. *Nature Reviews Neuroscience*, 25(2):111–130, 2024.

Appendix A. Biomarker Glossary Table

Table 2: Glossary of Biomarkers with Source, Units, and Interpretation

Abbrev.	**Full Name**	**Source Modality**	**Unit / Scale**	**Higher Values Indicate**
MMSE	Mini-Mental State Examination	Cognitive test	Score (0–30)	Less pathology (better global cognition)
ADAS	Alzheimer's Disease Assessment Scale – Cognitive	Cognitive test	Score (0–70)	More pathology (worse cognition)
AVLT-Sum	Auditory Verbal Learning Test – Sum Trials 1–5	Cognitive test	Score (0–75)	Less pathology (better memory encoding)
AB	Amyloid Beta ($A\beta_{1-42}$)	CSF	pg/mL	Less pathology (less amyloid deposition)
P-Tau	Phosphorylated Tau (e.g., p-Tau_{181})	CSF	pg/mL	More pathology (neurofibrillary tangle burden)
HIP-FCI	Hippocampal Functional Connectivity Index	Resting-state fMRI	Unitless	More pathology (abnormal hyperconnectivity)
PCC-FCI	Posterior Cingulate Cortex Functional Connectivity	Resting-state fMRI	Unitless	Less pathology (preserved DMN connectivity)
FUS-FCI	Fusiform Gyrus Functional Connectivity Index	Resting-state fMRI	Unitless	More pathology (compensatory hyperactivation)
HIP-GMI	Hippocampal Gray Matter Integrity	Structural MRI	Unitless	Less pathology (greater structural integrity)
FUS-GMI	Fusiform Gyrus Gray Matter Integrity	Structural MRI	Unitless	Less pathology (greater structural integrity)

Appendix B. SA-EBM Parameter Estimation Variants

B.1. Hard K-Means (Baseline)

Parameters are estimated once at initialization using K-Means clustering with conjugate priors and remain fixed throughout MCMC. This resembles the static parameter approach used in previous EBM implementations, but in a way that aligns more closely with the original EBM framework by explicitly fitting two separate states of the biomarker rather than a mixture distribution.

B.2. Conjugate Priors

Let $x_{n,j}$ denote the measurement of biomarker n for participant j. Both Conjugate Priors and Maximum Likelihood Estimation require hard assignments of x_{n_j} to either the pre-event or post-event cluster when estimating $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. If j is healthy, then $x_{n,j}$ is assigned to the pre-event cluster. Otherwise, the assignment is based on the stage posterior distribution:

$$P_{\text{pre-event}}(x_{n,j}) = \sum_{k_j \in \{1,2,\ldots,N\}} I(k_j < S(n)) P(k_j \mid \boldsymbol{X}_j, \boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\pi}) \tag{7}$$

$$P_{\text{post-event}}(x_{n,j}) = 1 - P_{\text{pre-event}}(x_{n,j}) \tag{8}$$

The measurement $x_{n,j}$ is assigned to the cluster with the larger probability. In the case where $P_{\text{pre-event}}(x_{n,j}) = P_{\text{post-event}}(x_{n,j})$, $x_{n,j}$ is assigned randomly with equal probability to either cluster.

Let $\boldsymbol{X}_{n,c}$ denote all the measurements of biomarker n in cluster c, where c represents either the pre-event or post-event cluster. The mean and variance of $\boldsymbol{X}_{n,c}$ are:

$$\bar{x} = \frac{1}{q} \sum_{i=1}^{q} \boldsymbol{X}_{(nc)_i} \tag{9}$$

$$s^2 = \frac{1}{q-1} \sum_{i=1}^{q} \left(\boldsymbol{X}_{(nc)_i} - \bar{x} \right)^2 \tag{10}$$

where q is the size of $\boldsymbol{X}_{n,c}$. Assuming $\boldsymbol{X}_{n,c}$ follow Gaussian distributions, we employ Normal-Inverse-Gamma priors where parameters from the previous iteration serve as priors for the current update. Given observations $\boldsymbol{X}_{n,c}$ and prior hyperparameters (m_0, n_0, s_0^2, v_0), the posterior parameters are:

$$m_n = \frac{n_0 m_0 + q\bar{x}}{n_0 + q} \tag{11}$$

$$n_n = n_0 + q \tag{12}$$

$$v_n = v_0 + q \tag{13}$$

$$s_n^2 = \frac{1}{v_n} \left[(q-1)s^2 + v_0 s_0^2 + \frac{n_0 q}{n_n} (\bar{x} - m_0)^2 \right] \tag{14}$$

where m_0, and s_0^2 are prior estimates of mean and variance. m_n, and s_n^2 are the resulting posterior estimates. n_0 and n_n are strengths of belief in m_0 and m_n, respectively. v_0 and v_n represent degrees of freedom, influencing the certainty of s_0^2 and s_n^2. Initially, $n_0 = v_0 = 1$, in the spirit of weakly informative priors (Gelman et al., 2017). If c is the post-event cluster, then $\mu_{n,\theta} = m_n, \sigma_{n,\theta}^2 = s_n^2$.

B.3. Maximum Likelihood Estimation (MLE)

Parameters $\boldsymbol{\theta}_n$ and $\boldsymbol{\phi}_n$ are updated using standard MLE after assignment of all $x_{n,c}$. If c is the post-event cluster, then $\mu_{n,\theta} = \bar{x}, \sigma_{n,\theta}^2 = s^2$.

B.4. Expectation-Maximization (EM)

Instead of hard assignments, measurements $x_{n,c}$ are soft-assigned based on stage posteriors. For example, $\mu_{n,\theta}$ and $\sigma_{n,\theta}^2$ are estimated as:

$$\mu_{n,\theta} = \frac{\sum_{j=1}^{J} P_{\text{post-event}}(x_{n,j}) \cdot x_{n,j}}{\max\left(10^{-9}, \sum_{j=1}^{J} P_{\text{post-event}}(x_{n,j})\right)} \tag{15}$$

$$\sigma_{n,\theta}^2 = \frac{\sum_{j=1}^{J} P_{\text{post-event}}(x_{n,j}) \cdot (x_{n,j} - \mu_{n,\theta})^2}{\max\left(10^{-9}, \sum_{j=1}^{J} P_{\text{post-event}}(x_{n,j})\right)} \tag{16}$$

where $P_{\text{pre-event}}$ and $P_{\text{post-event}}$ can be obtained through Equation 7 and 8.

B.5. Kernel Density Estimation (KDE)

We use Gaussian kernels with Scott's bandwidth selection rule. Weights are updated using the same soft-assignment approach as EM. More specifically:

$$h = \sigma_w \cdot n_{\text{eff}}^{-1/5}$$

where

$$\mu_w = \frac{\sum w_i x_i}{\sum w_i}$$

$$\sigma_w^2 = \frac{\sum w_i (x_i - \mu_w)^2}{\sum w_i}$$

$$n_{\text{eff}} = \frac{1}{\sum w_i^2}$$

A lower bound (10^{-12}) of σ_w is applied to avoid division by zero.
We used a Gaussian kernel for density estimation:

$$\hat{f}_h(x) = \sum_{i=1}^{n} w_i \frac{1}{\sqrt{2\pi h^2}} e^{-\frac{1}{2}\left(\frac{x - x_i}{h}\right)^2}$$

where

- h: the selected bandwidth using Scott's rule detailed above.
- x: an arbitrary point for density estimation
- $\hat{f}_h(x)$: the estimated probability density for data point x.
- n: the number of measurements in the dataset.
- x_i: the i-th observed biomarker measurement in the dataset.
- w_i: normalized weights ($\sum_{i=1}^{n} w_i = 1$), representing the relative contribution of each observed data point.

Appendix C. Biomarker Parameters and Non-Normal Distribution Parameter Details

Table 3: Biomarker Parameters and Non-Normal Sampling Specifications

Biomarker	θ_{mean}	θ_{std}	ϕ_{mean}	ϕ_{std}	**Non-Normal Components (Per Code Implementation)**
MMSE	22	2.67	28	0.67	1. Triangular(left=$\mu - 2\sigma$, mode=$\mu - 1.5\sigma$, right=μ) 2. $\mathcal{N}(\mu + \sigma,\ (0.3\sigma)^2)$ 3. Exp$(0.7\sigma) + (\mu - 0.5\sigma)$ - Equal 3-way split & combined
ADAS	20	4.00	6	1.33	Same component structure as MMSE
AB	150	16.67	250	50.00	1. Pareto$(1.5) \times \sigma + (\mu - 2\sigma)$ 2. $\mathcal{U}(\mu - 1.5\sigma,\ \mu + 1.5\sigma)$ 3. Logistic$(\mu,\ \sigma)$ - Equal 3-way split & combined
P-Tau	50	33.33	25	16.67	Same component structure as AB
HIP-FCI	5	6.67	-5	1.67	1. Beta$(0.5, 0.5) \times 4\sigma + (\mu - 2\sigma)$ 2. Exp$(0.4\sigma) \times$ sign$(Bernoulli(0.5))$ 3. $\mathcal{N}(\mu,\ (0.5\sigma)^2) + \{0, 2\sigma\}$ spikes - Equal 3-way split & combined
HIP-GMI	0.3	0.33	0.4	0.23	Same component structure as HIP-FCI
AVLT-Sum	20	6.67	40	15.00	1. Gamma$(2, 0.5\sigma) + (\mu - \sigma)$ 2. Weibull$(1.0) \times \sigma + (\mu - \sigma)$ 3. $\mathcal{N}(\mu,\ (0.5\sigma)^2) \pm \sigma$ - Equal 3-way split & combined
PCC-FCI	5	3.33	12	4.00	Same component structure as AVLT-Sum
FUS-GMI	0.5	0.07	0.6	0.07	Cauchy$(\mu,\ \sigma) + \mathcal{N}(0,\ (0.2\sigma)^2)$ Clipped to $[\mu - 4\sigma,\ \mu + 4\sigma]$
FUS-FCI	20	6.00	10	3.33	10%: $\mathcal{N}(\mu,\ (0.2\sigma)^2)$ 90%: Logistic$(\mu + \sigma,\ 2\sigma)$

Implementation Notes:

- μ & σ use θ parameters for affected (pathological) and ϕ for nonaffected (intact).
- For non-normal components, **After sampling, all values are perturbed by additional noise** $\mathcal{N}(0,\ (0.2\sigma)^2)$ **and clipped to** $[\mu - 5\sigma,\ \mu + 5\sigma]$.

Appendix D. Experimental Specifications

Table 4: Complete Experimental Specifications with Defined Notation

Exp	Model	Event Time	Stage Distribution (for $k_j > 0$)	Biomarker Measurements
1	EBM	Uniform permutation (ordinal)	Dirichlet-Multinomial ($\alpha = [0.40, 1.09, 2.31, 3.81, 4.89, 4.89, 3.81, 2.31, 1.09, 0.40]$)	Pre-event: $\mathcal{N}(\mu_n^{\text{pre}}, (\sigma_n^{\text{pre}})^2)$ Post-event: $\mathcal{N}(\mu_n^{\text{post}}, (\sigma_n^{\text{post}})^2)$
2	EBM	Same as Exp1	Same as Exp1	Biomarker-specific mixtures (see Table 3)
3	EBM	Same as Exp1	Dirichlet-Multinomial ($\alpha_i = 100$)	Same normal structure as Exp1
4	EBM	Same as Exp1	Same as Exp3	Same mixtures as Exp2
5	Sigmoid	Same as Exp1	$\text{Beta}(1,1) \times N$	Post-event: Pre-event + $\frac{(-1)^{I_n} R_n}{1 + e^{-\rho_n (k_j - S_n)}}$ Pre-event: $\mathcal{N}(\mu_n^{\text{pre}}, (\sigma_n^{\text{pre}})^2)$
6	Sigmoid	Same as Exp1	$\text{Beta}(5,2) \times N$	Same as Exp5
7	Sigmoid	$\text{Beta}(2,2) \times N$	$\text{Beta}(1,1) \times N$	Post-event: Pre-event + $\frac{(-1)^{I_n} R_n}{1 + e^{-\rho_n (k_j - \xi_n)}}$ Pre-event: $\mathcal{N}(\mu_n^{\text{pre}}, (\sigma_n^{\text{pre}})^2)$
8	Sigmoid	Same as Exp7	$\text{Beta}(5,2) \times N$	Same as Exp7
9	Sigmoid	$\text{clip}\big(\text{Beta}(2,2) \times N + \mathcal{N}(0,\ 0.05 \cdot N),\ 0,\ N\big)$	Same as Exp8	Same as Exp7

Notation Clarifications:

- $\mu_n^{\text{pre}}, \sigma_n^{\text{pre}}$: Pre-event parameters for biomarker n
- $\mu_n^{\text{post}}, \sigma_n^{\text{post}}$: Post-event parameters
- $R_n = \mu_n^{\text{post}} - \mu_n^{\text{pre}}$: Biomarker dynamic range
- $\text{clip}(x, a, b) = \min(\max(x, a), b)$
- $\times N$: Scales value to the interval $(0,\ N]$

Appendix E. Distributions Used in Experiments

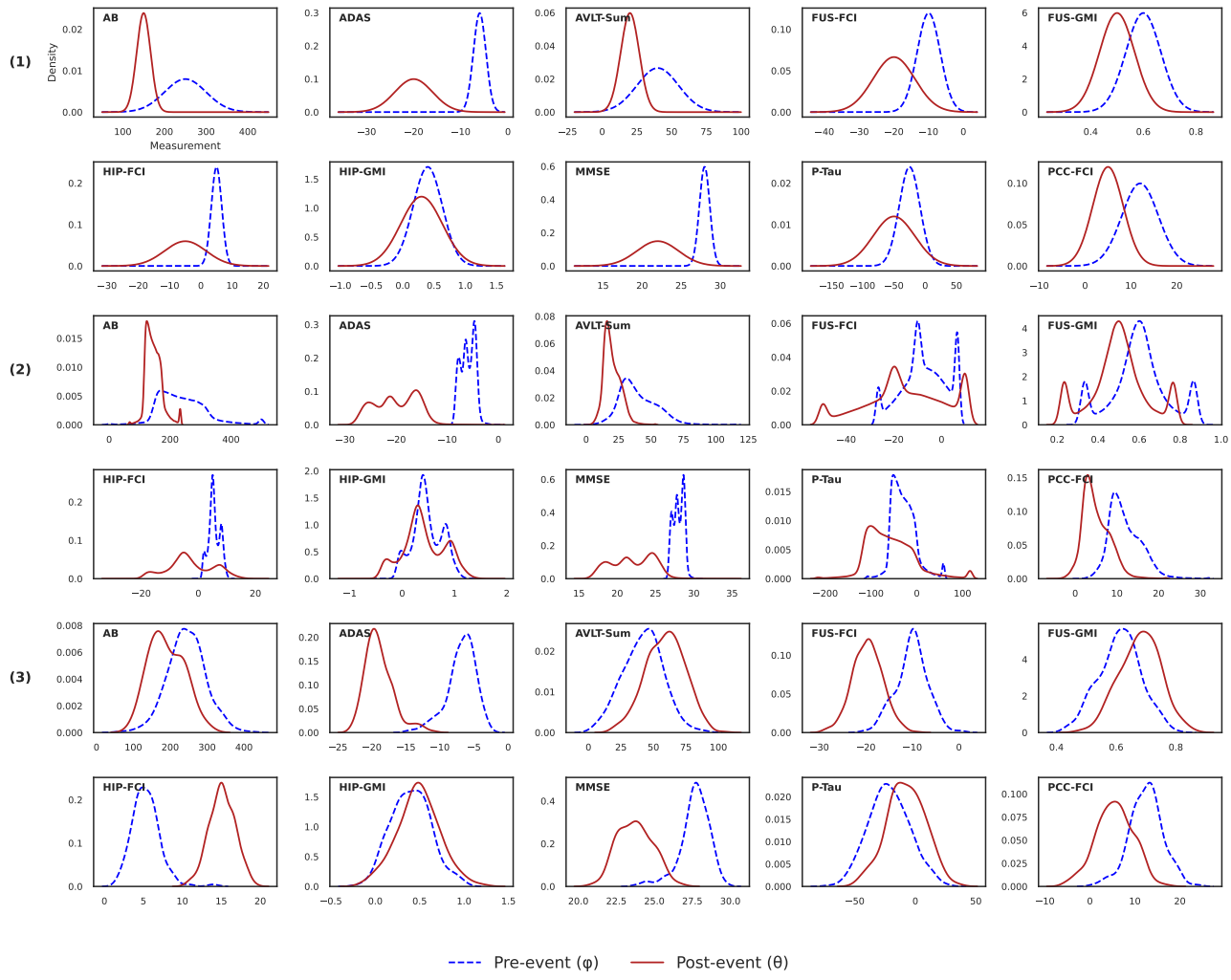

Figure 4: **(1) Theoretical normal distributions; (2) Theoretical non-normal distributions; (3) Empirical distributions in one dataset of experiment 9.**

Appendix F. Detailed Results of Synthetic Experiments

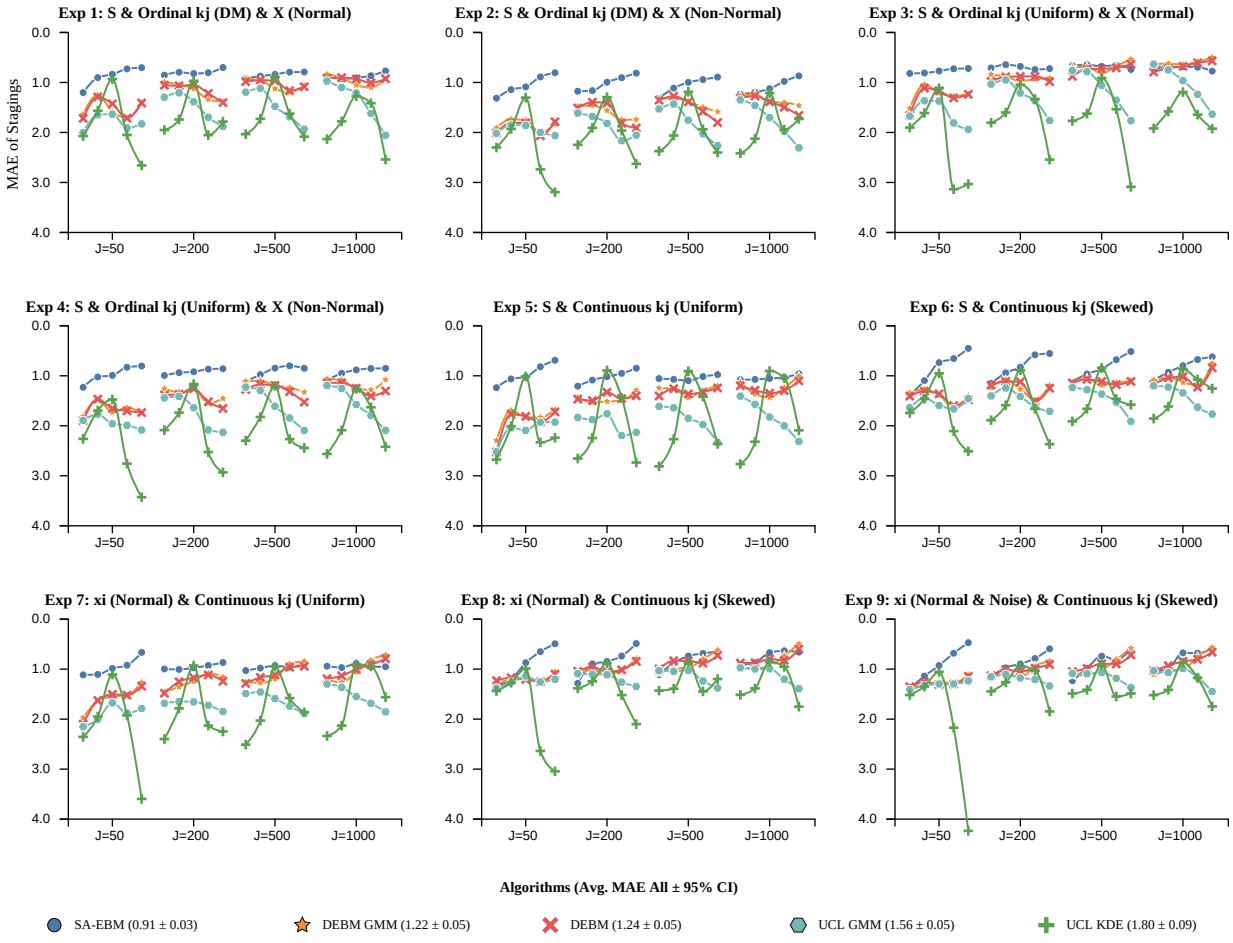

Figure 5: **Mean average errors (±95% CI)** for staging accuracy across nine synthetic experiments. Results are organized in the same way as in Figure 2. SA-EBM outperforms static methods in staging accuracy.

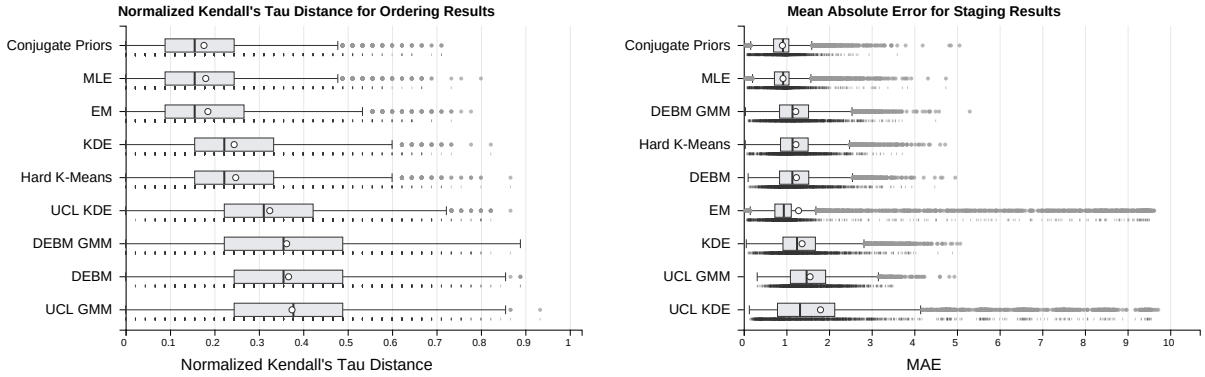

Figure 6: **Aggregated algorithm performance**: The left panel shows ordering accuracy (normalized Kendall's Tau distance) and the right panel displays staging accuracy. Box plots represent the performance of each algorithm across all 9,000 datasets. Representative samples of individual data points from each corresponding result are visualized underneath. Algorithms are sorted by the average performance represented by the open circle. The five variants of SA-EBM are: Conjugate Priors, MLE, EM, KDE and Hard K-Means.

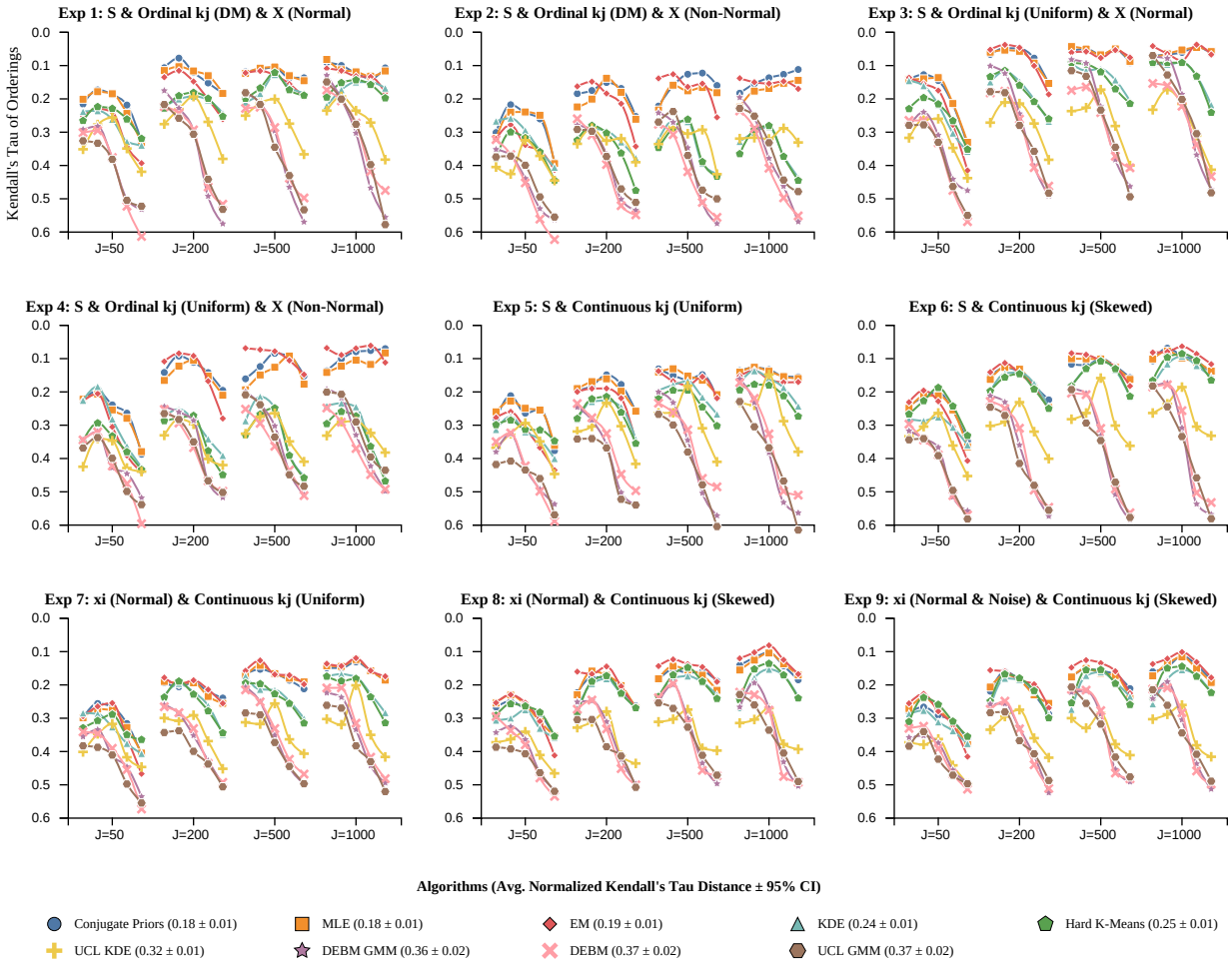

Figure 7: **Average Kendall's Tau distance values (±95% CI)** of all algorithms across nine synthetic experiments. Each panel represents a different experimental configuration with varying data generation models, stage distributions, and biomarker distributions. The x-axis within each panel shows different participant sizes ($J = 50, 200, 500, 1000$). Within each participant size are different healthy ratios (r), i.e., the percentage of healthy participants among all subjects. From left to right are $r = 0.1, 0.25, 0.5, 0.75, 0.9$. The y-axis shows Kendall's Tau value (higher is better). Data points represent mean performance across 50 variants of the same experimental configuration, sample size, and healthy ratio. SA-EBM variants (Conjugate Priors, MLE, and EM) consistently outperform static and baseline methods. Performance generally improves with increasing sample size, while fixed-parameter methods degrade under high healthy ratios and non-Gaussian data.

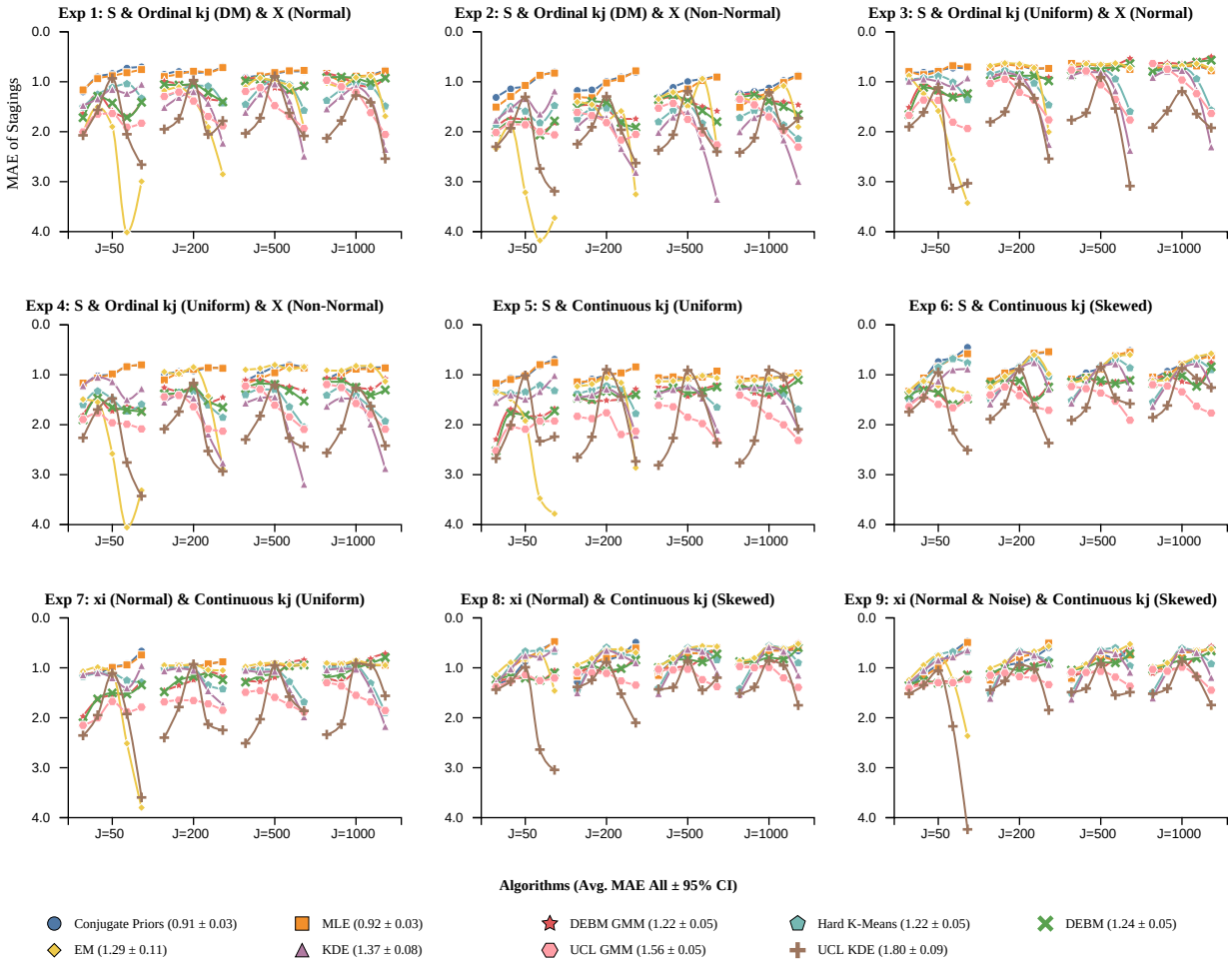

Figure 8: **Mean Average Errors (±95% CI)** for staging accuracy of all algorithms across nine synthetic experiments. Results are organized in the same way as in Figure 2. SA-EBM outperforms static methods in staging accuracy.

Appendix G. Sigmoid Transitions

Without considering the noise from $\mathcal{N}(\mu_{n,\phi}, \sigma^2_{n,\phi})$, the ideal development of biomarkers is modeled as:

$$X_{n,j} = \mu_{n,\phi} + \frac{(-1)^{I_n} R_n}{1 + \exp\left(-\rho_n (k_j - \xi_n)\right)}$$

For visualization purposes, we normalize each biomarker's trajectory using min-max normalization:

$$\text{Norm}(X_{n,j}) = \frac{X_{n,j} - \min_{k_j} X_{n,j}}{\max_{k_j} X_{nj} - \min_{k_j} X_{n,j}}$$

where the minimum and maximum are taken over all disease stages k_j.

Note that each curve represents the normalized ideal trajectory of a biomarker across disease stages.

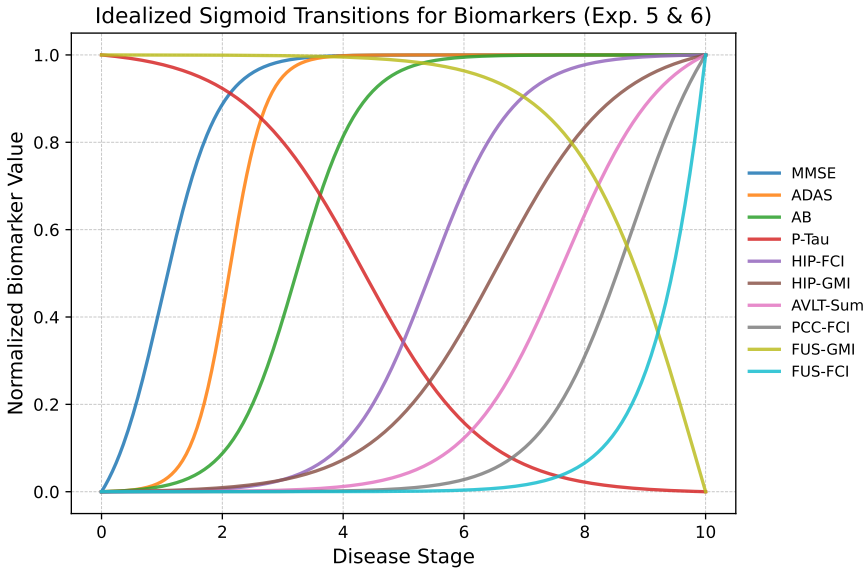

Figure 9: **Normalized sigmoid progression of biomarkers for experiments 5 & 6**

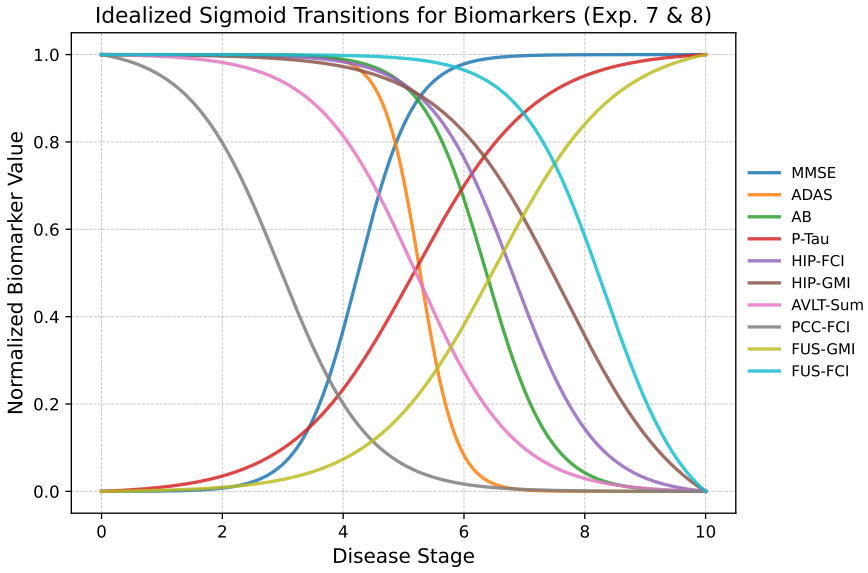

Figure 10: **Normalized sigmoid progression of biomarkers for experiments 7 & 8**

Appendix H. ADNI Data Processing Pipeline

We used the `adnimerge` table, which consolidates data from the Alzheimer's Disease Cooperative Study (ADCS) data system. The version we accessed was last updated on September 7, 2023. We processed and filtered data using the following steps:

- Included only participants' baseline visits, identified by `VISCODE = bl`.
- Included only participants whose baseline diagnosis was Control (CN), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), or Alzheimer' Disease (AD).
- Selected twelve biomarkers commonly reported in previous studies, e.g., Cs et al. (2025), Young et al. (2014), and Archetti et al. (2019). These biomarkers include cognitive assessments (MMSE, ADAS13, RAVLT immediate), cerebrospinal fluid (CSF) markers associated with tau and amyloid pathology (PTAU, TAU, ABETA), and structural MRI-derived volumetric measures of specific brain regions (Ventricles, Whole-Brain, MidTemp, Fusiform, Entorhinal, Hippocampus). We excluded participants with missing values for any of selected biomarkers.
- Removed Duplicate observations.

Appendix I. ADNI Participants Age Distribution

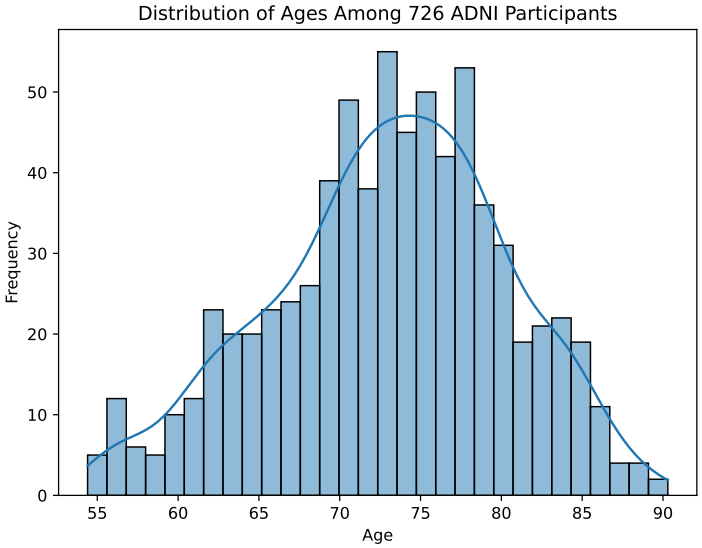

Figure 11: **Age distribution of ADNI participants**

Appendix J. SA-EBM Trace Plots on ADNI

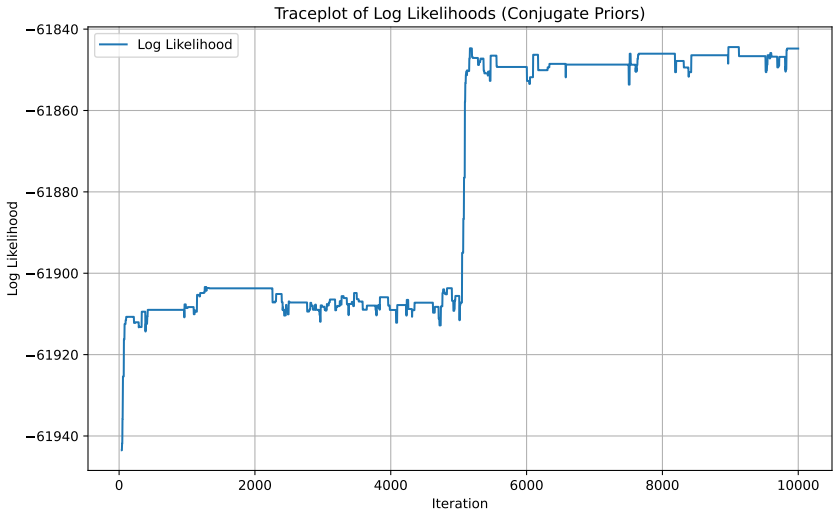

Figure 12: **Traceplot of log data likelihood for Conjugate Priors on ADNI data**

Appendix K. Ordering Results for ADNI

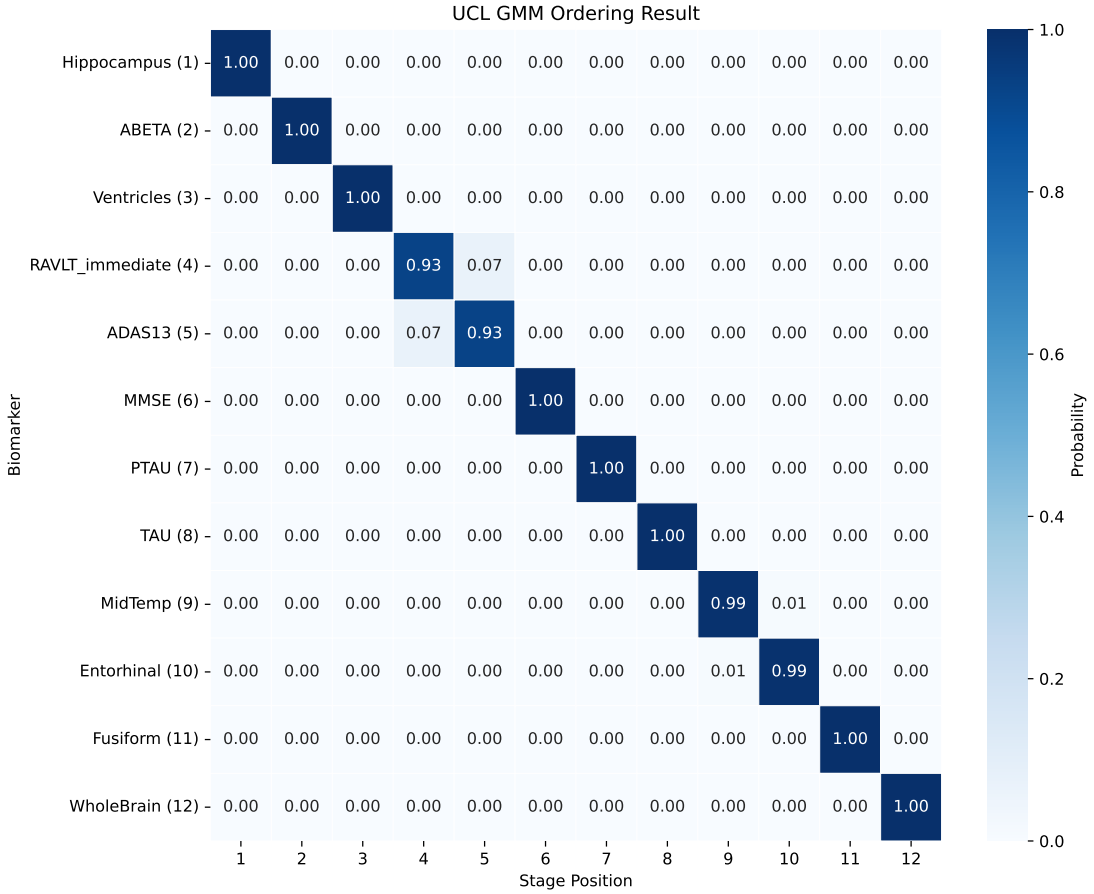

Figure 13: **Ordering result with UCL GMM on ADNI data**: The heatmap shows uncertainties for 10,000 MCMC iterations. The number inside the parenthesis in the Y-axis indicates the result according to the ordering associated with the largest log data likelihood.

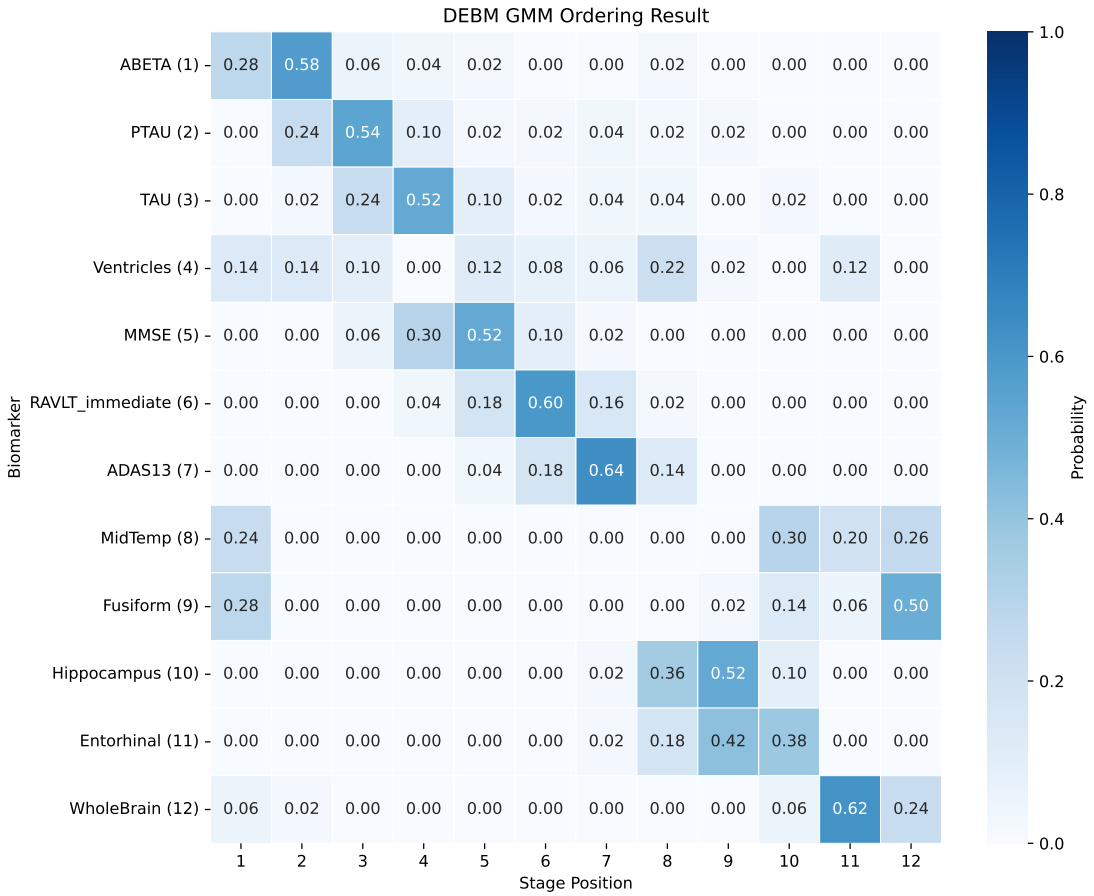

Figure 14: **Ordering result with DEBM GMM on ADNI data**: The heatmap shows uncertainties for 50 Bootstraps. The number inside the parenthesis in the Y-axis indicates the result according to the "MeanCentralOrdering" as generated by DEBM (Venkatraghavan et al., 2019).

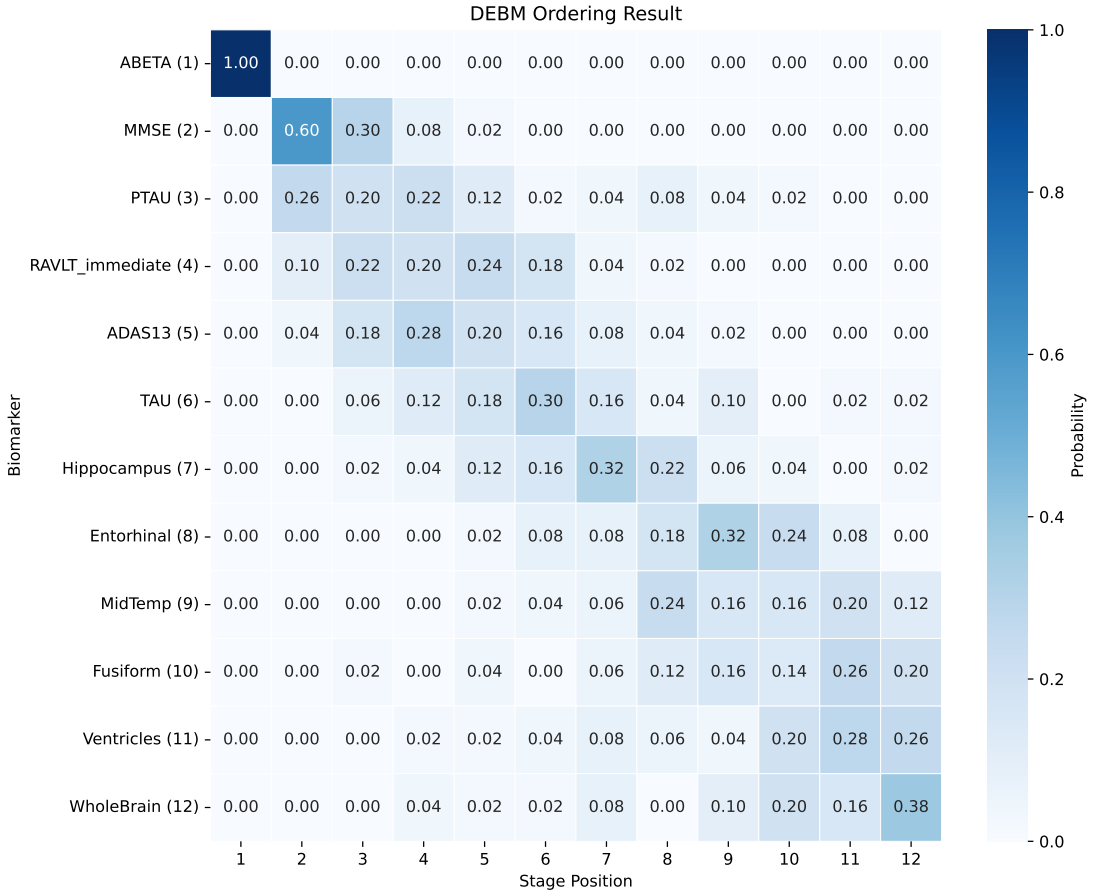

Figure 15: **Ordering result with DEBM on ADNI data**: The heatmap shows uncertainties for 50 Bootstraps. The number inside the parenthesis in the Y-axis indicates the result according to the "MeanCentralOrdering" as generated by DEBM (Venkatraghavan et al., 2019).

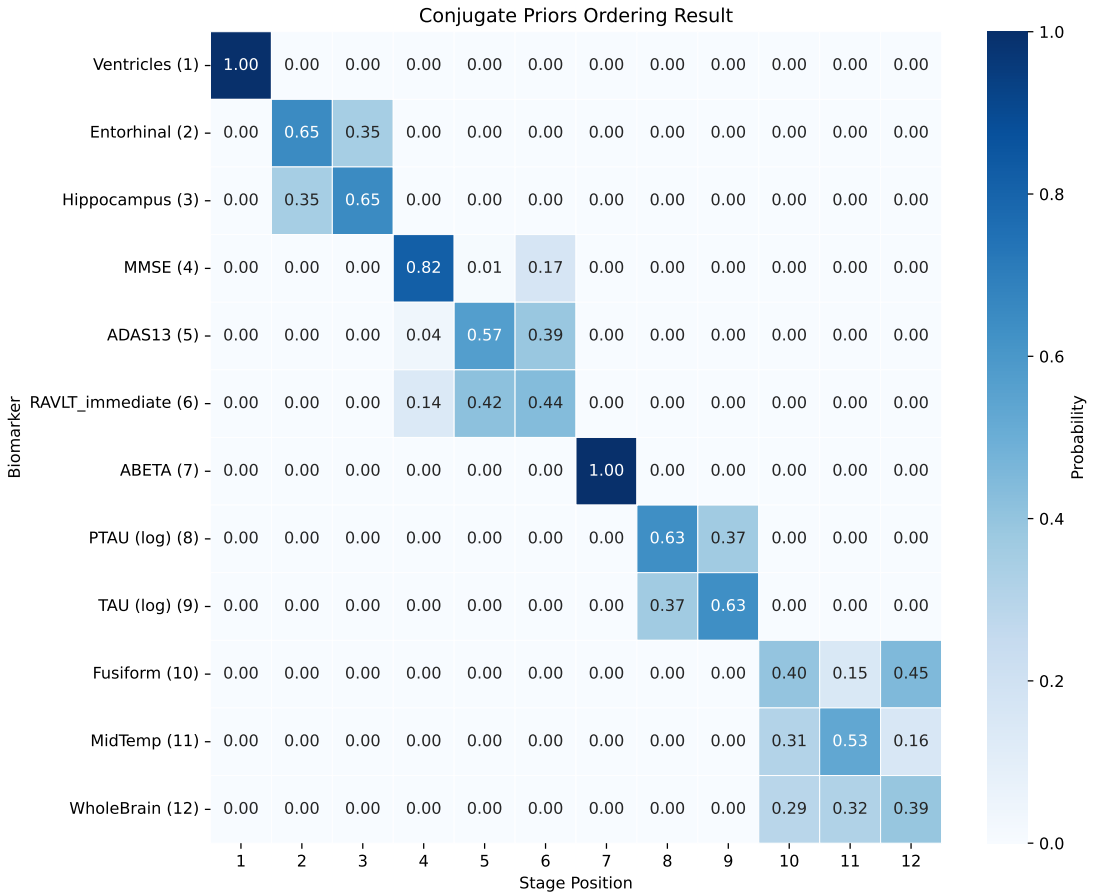

Figure 16: **Ordering result with Conjugate Priors on ADNI data (Log Transformation on TAU and PTAU measurements)**: The heatmap shows uncertainties for 10,000 MCMC iterations with 500 burn-in and no thinning. The number inside the parenthesis in the Y-axis indicates the result according to the ordering associated with the largest log data likelihood. Each cell indicates the probability of each biomarker getting affected in a specific stage, according to the results from the last 9,500 MCMC iterations.

Appendix L. Staging Results for ADNI

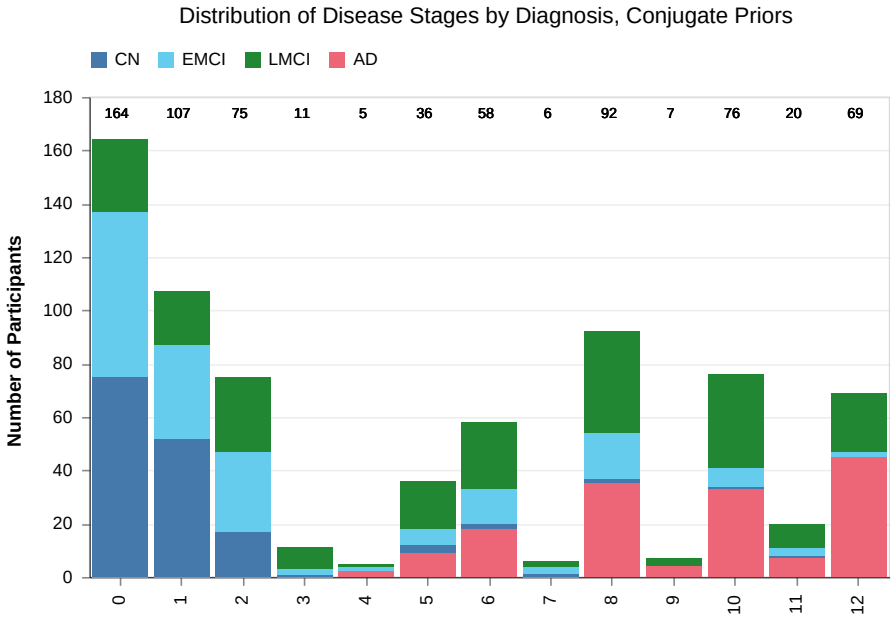

Figure 17: **Estimated distribution of disease stages by diagnosis, using Conjugate Priors**: Note that for the staging task, we ignored the known diagnosis label, and let each algorithm infer the staging solely based on each participant's biomarker measurements. It is clear that disease stages are unevenly distributed. EMCI participants were the majority in early stages, but LMCI and AD became the majority later on, validating our SA-EBM.

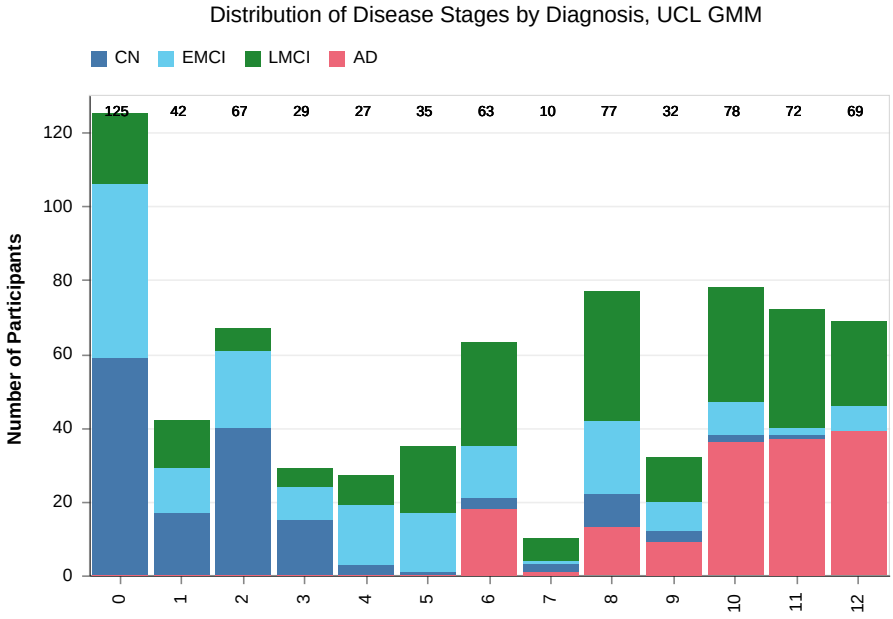

Figure 18: **Estimated distribution of disease stages by diagnosis, using UCL GMM**

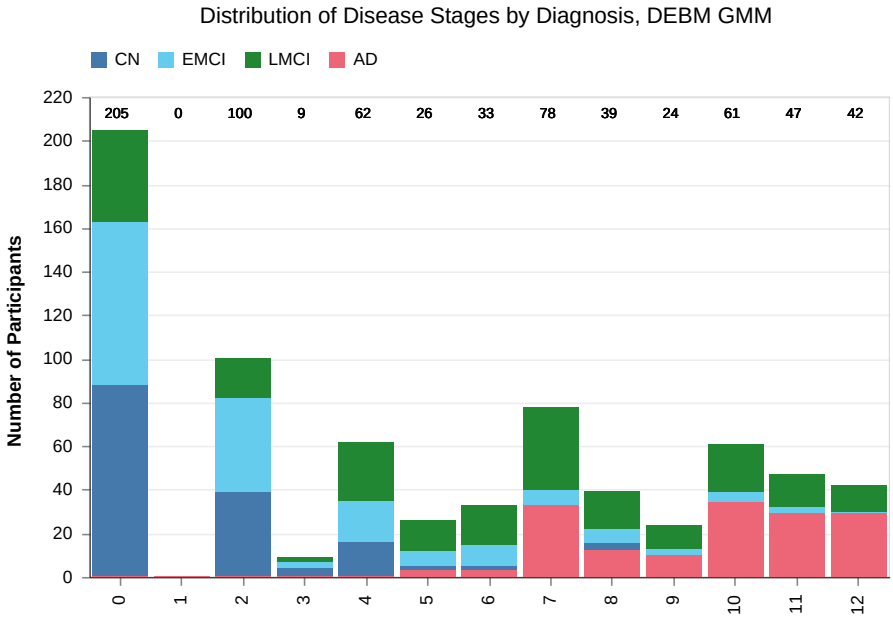

Figure 19: **Estimated distribution of disease stages by diagnosis, using DEBM GMM**

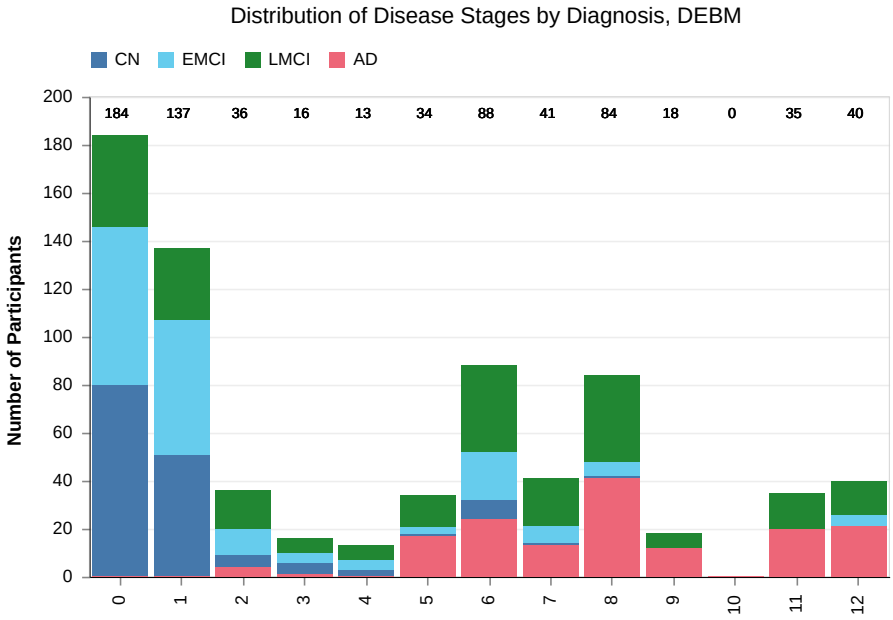

Figure 20: **Estimated distribution of disease stages by diagnosis, using DEBM**

Appendix M. Runtime Comparison

Table 5: Average runtime (in minutes) for EBM algorithms

Algorithm	J			
	50	200	500	1000
UCL GMM	0.05	0.06	0.07	0.09
DEBM	0.06	0.07	0.07	0.07
UCL KDE	0.04	0.05	0.07	0.12
DEBM GMM	0.12	0.14	0.15	0.17
Hard K-Means	1.35	4.85	11.96	23.35
EM	2.84	8.97	21.33	41.17
MLE	3.12	9.94	23.70	46.25
Conjugate Priors	3.14	9.96	23.74	46.32
KDE	4.07	18.38	67.34	208.81

Appendix N. Data and Code Availability

- Package: https://pypi.org/project/pysaebm
- Package Source Code: https://github.com/jpcca/pysaebm
- Reproducible Code for Experiments in this study: https://github.com/hongtaoh/saebm

Appendix O. ADNI Information

O.1. ADNI Investigators

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

O.2. ADNI Acknowledgment

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