Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling

Vikram Venkatraghavan [a,*], Esther E. Bron [a], Wiro J. Niessen [a,b], Stefan Klein [a], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] Biomedical Imaging Group Rotterdam, Departments of Medical Informatics & Radiology, Erasmus MC, University Medical Center Rotterdam, the Netherlands
[b] Quantitative Imaging Group, Dept. of Imaging Physics, Faculty of Applied Sciences, Delft University of Technology, Delft, the Netherlands

ABSTRACT

Alzheimer's Disease (AD) is characterized by a cascade of biomarkers becoming abnormal, the pathophysiology of which is very complex and largely unknown. Event-based modeling (EBM) is a data-driven technique to estimate the sequence in which biomarkers for a disease become abnormal based on cross-sectional data. It can help in understanding the dynamics of disease progression and facilitate early diagnosis and prognosis by staging patients. In this work we propose a novel discriminative approach to EBM, which is shown to be more accurate than existing state-of-the-art EBM methods. The method first estimates for each subject an approximate ordering of events. Subsequently, the central ordering over all subjects is estimated by fitting a generalized Mallows model to these approximate subject-specific orderings based on a novel probabilistic Kendall's Tau distance. We also introduce the concept of relative distance between events which helps in creating a disease progression timeline. Subsequently, we propose a method to stage subjects by placing them on the estimated disease progression timeline. We evaluated the proposed method on Alzheimer's Disease Neuroimaging Initiative (ADNI) data and compared the results with existing state-of-the-art EBM methods. We also performed extensive experiments on synthetic data simulating the progression of Alzheimer's disease. The event orderings obtained on ADNI data seem plausible and are in agreement with the current understanding of progression of AD. The proposed patient staging algorithm performed consistently better than that of state-of-the-art EBM methods. Event orderings obtained in simulation experiments were more accurate than those of other EBM methods and the estimated disease progression timeline was observed to correlate with the timeline of actual disease progression. The results of these experiments are encouraging and suggest that discriminative EBM is a promising approach to disease progression modeling.

1. Introduction

Dementia is considered a major global health problem as the number of people living with dementia was estimated to be about 46.8 million in 2015. It is expected to increase to 131.5 million in 2050 (Prince et al., 2015). Alzheimer's Disease (AD) is the most common form of dementia. There is a gradual shift in the definition of AD from it being a clinical-pathologic entity (based on clinical symptoms), to a biological one based on neuropathologic change (change of imaging and non-imaging biomarkers from normal to abnormal) (Jack et al., 2018). The latter definition is more useful for understanding the mechanisms of disease progression.

Preventive and supportive therapy for patients at risk of developing dementia due to AD could improve their quality of life and reduce costs related to care and lifestyle changes. To identify the at-risk individuals as well as monitor the effectiveness of these preventive and supportive therapies, methods for accurate patient staging (estimating the disease severity in each individual) are needed. To enable accurate patient staging in an objective and quantitative way, it is important to understand how the different imaging and non-imaging biomarkers progress after disease onset.

Longitudinal models of disease progression, such as Jedynak et al. (2012), reconstruct long-term biomarker trajectories using short term data. Donohue et al. (2014) estimate these trajectories based on
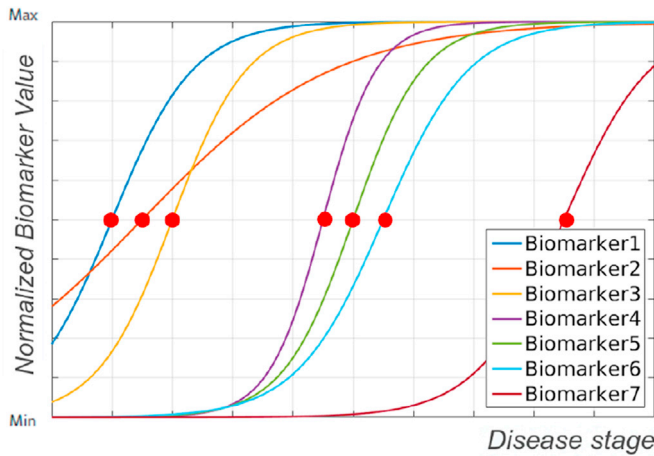
Fig. 1. Illustration of the output expected in an EBM. The biomarker trajectories shown here are hypothetical trajectories representing a change of biomarker value from normal state. The dots on these trajectories are biomarker events as defined in an EBM. Output of an EBM is the ordering of such events.

self-modeling regression, whereas Cox regression was used in Sabuncu et al. (2014). Rather than focussing only on a mean trajectory for the entire population, Schmidt-Richberg et al. (2016) estimate percentile curves based on quantile regression. Li et al. (2017); Schiratti et al. (2015) estimate subject-specific trajectories using a mixed model. Lorenzi et al. (2017) provide a probabilistic estimate of biomarker trajectories. While such models are useful for understanding disease progression, their utility in identifying at-risk individuals is restricted. This is due to the fact that selecting a cohort of at-risk individuals for clinical trials based on a longitudinal dataset is not feasible (Marinescu et al., 2018). The utility of these models in studying other forms of dementia is also restricted because longitudinal data in large groups of patients is often scarce.

To circumvent this problem, methods to infer the order in which biomarkers become abnormal during disease progression using cross-sectional data have been proposed (Fonteijn et al., 2012; Huang and Alexander, 2012; Iturria-Medina et al., 2016). The model used in Iturria-Medina et al. (2016) relies on stratification of patients into several subgroups based on symptomatic staging, for inferring the aforementioned ordering. However, the problem with using symptomatic staging is that it is very coarse and qualitative. The models used in Fonteijn et al. (2012); Huang and Alexander (2012) are variants of Event-Based Models (EBM). EBM algorithms neither rely on symptomatic staging nor on the presence of longitudinal data for inferring the temporal ordering of events, where an event is defined by a biomarker becoming abnormal. Fig. 1 shows these biomarker events on hypothetical trajectories as expected in a typical neuropathologic change.

An important assumption made in Fonteijn et al. (2012) is that the ordering of events is common for all the subjects in a dataset. AD is known to be a heterogeneous disease with multiple disease subtypes. The assumptions in Fonteijn's EBM may therefore be too restrictive. The assumptions in Huang's EBM on the other hand are more realistic, as they do assume that the disease is heterogeneous. However the algorithm does not scale well to a large number of biomarkers (Venkatraghavan et al., 2017).

To make EBM more scalable to large number of biomarkers and subjects, as well as make it robust to variations in ordering, we propose a novel approach to EBM, discriminative event-based model (DEBM), for

estimating the ordering of events.[2] We also introduce the concept of relative distance between events which helps in creating a disease progression timeline. Subsequently, we propose a method to stage subjects by placing them on the estimated disease progression timeline. The other contributions of this paper include an optimization technique for Gaussian mixture modeling that helps in accurate estimation of event ordering in DEBM as well as improving the accuracies of other EBMs, and a novel probabilistic distance metric between event orderings (probabilistic Kendall's Tau).

The remainder of the paper is organized as follows: An introduction to the existing EBM models is given in Section 2. In Section 3, we propose our novel method for estimating central ordering of events. We perform extensive sets of experiments on ADNI data as well as on simulation data, the details of which are in Section 4. Section 5 summarizes the results of the experiments. Section 6 discusses the implications of these findings followed by concluding remarks in Section 7.

2. Event-Based Models

EBM assumes monotonic increase or decrease of biomarker values with increase in disease severity (with the exception of measurement noise). It considers disease progression as a series of events, where each event corresponds to a new biomarker becoming abnormal. Fonteijn's EBM (Fonteijn et al., 2012) finds the ordering of events (S) such that the likelihood that a dataset was generated from subjects following this event ordering is maximized. S is a set of integer indices of biomarkers, which represents the order in which they become abnormal. Thus, disease progression is defined by $\{E_{S(1)}, E_{S(2)}, ..., E_{S(N)}\}$, where N is the number of biomarkers per subject in the dataset and $E_{S(i)}$ is the i-th event that is associated with biomarker $S(i)$ becoming abnormal.

In a cross-sectional dataset (X) of M subjects, X_j denotes a measurement of biomarkers for subject $j \in [1, M]$, consisting of N scalar biomarker values $x_{j,i}$. Probabilistic formulation of an EBM, as proposed in Fonteijn et al. (2012), can be given by $argmax_S(p(S|X))$, where can be written using Bayes' rule as:

$$p(S|X) = \frac{p(S)p(X|S)}{p(X)} \qquad (1)$$

An important assumption in Fonteijn et al. (2012) is that $p(S)$ is uniformly distributed. This makes inferring S, equivalent to the maximum likelihood problem of maximizing $p(X|S)$[3]. This can be further written in terms of X_j as follows:

$$p(X|S) = \prod_{j=1}^{M} p(X_j|S) \qquad (2)$$

where $p(X_j|S)$ can be written as:

$$p(X_j|S) = \sum_{k=0}^{N} p(k|S)p(X_j|k, S) \qquad (3)$$

where $p(k|S)$ is the prior probability of a subject being at position k of the event ordering, which is assumed to be equal for each position. The k which maximizes $p(X_j|S)$ denotes subject j's disease stage. This method of identifying disease severity for a subject results in discrete set of stages, where the number of stages is one more than the number of biomarkers used for creating the model. $p(X_j|k, S)$ can be expressed as:

---

[3] Fonteijn's EBM uses Markov Chain Monte Carlo (MCMC) sampling to estimate the posterior distribution $P(S|X)$. Average position of events in all the MCMC samples was used as a way for selecting the mean ordering by Fonteijn et al. (2012) whereas further extensions of the work such as Young et al. (2014) prefer the maximum likelihood solution.
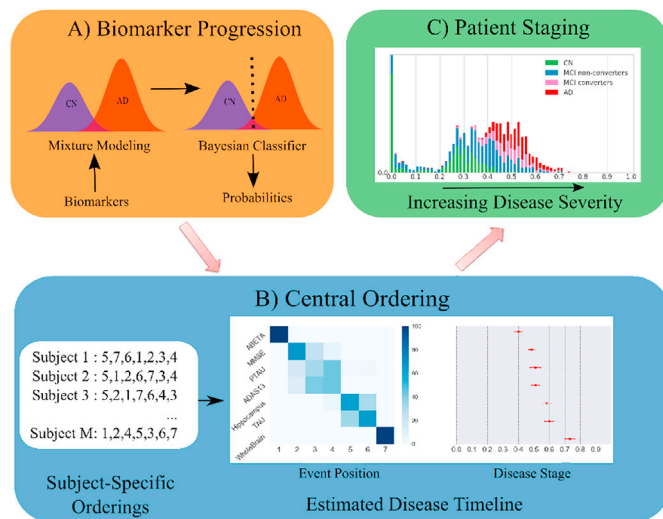
Fig. 2. Overview of the steps in DEBM. A) Biomarkers measured from different subjects are converted to probabilities of abnormality for individual biomarkers. This is done by estimating normal and abnormal distributions using Gaussian mixture modeling before classifying individual biomarkers using a Bayesian classifier. B) Subject-specific orderings of biomarker abnormalities are inferred from these probabilities which are then used to estimate the central ordering and for creating the disease progression timeline. C) This is then used to stage subjects based on disease severity.

$$p(X_j|k, S) = \prod_{i=1}^{k} p(x_{j,S(i)}|E_{S(i)}) \times \prod_{i=k+1}^{N} p(x_{j,S(i)}|\neg E_{S(i)}) \quad (4)$$

where $p(x_{j,S(i)}|E_{S(i)})$ is the likelihood of observing $x_{j,S(i)}$ in subject j, conditioned on event i having already occurred. $p(x_{j,S(i)}|\neg E_{S(i)})$, on the other hand, computes a similar likelihood, given that event i has not occurred.

With the assumption that all the biomarkers in the control population are normal and that the biomarker values follow a Gaussian distribution, $p(x_{j,S(i)}|\neg E_{S(i)})$ is computed. Abnormal biomarker values in the patient population are assumed to follow a uniform distribution but not all biomarkers of a patient could be assumed to be abnormal. For this reason, the likelihoods were obtained using a mixture model of a Gaussian and a uniform distribution, where only the parameters of the uniform distribution were allowed to be optimized.

This method was modified in Young et al. (2014) to estimate the optimal ordering in a sporadic AD dataset with significant proportions of controls expected to have presymptomatic AD (Schott et al., 2010). A Gaussian distribution was used to describe both the control and patient population, and the mixture model allowed for optimization of parameters for the Gaussians describing both control and patient population. The Gaussian mixture model was also used to incorporate more subjects from the dataset with clinical diagnosis of mild cognitive impairment (MCI).

After obtaining the central ordering S which maximizes the likelihood $p(X|S)$, staging of patients is done by finding a disease stage k for subject j, such that $p(X_j|k, S)$ is maximized.

The assumption that subjects follow a unique event ordering was relaxed by Huang and Alexander (2012), who estimate a distribution of event orderings with a central event ordering (S) and a spread (ϕ) as per a generalized Mallows model (Fligner and Verducci, 1988) using an expectation maximization algorithm. The E-step estimates the likelihood of patients' biomarker value measurements following subject-specific event order s_j, given S and ϕ. In the M-step, S and ϕ are estimated based on s_j estimated in the E-step. This is done iteratively to maximize the likelihood of generation of patients' data based on S and ϕ. Patient staging in Huang's EBM is also a maximum likelihood estimate, but

unlike Fonteijn's EBM, the staging is done on the subject-specific event ordering s_j.

In both Fonteijn's and Huang's EBM, relative distances between events, that can be observed in Fig. 1, are not captured.[4] Some events can be closer to each other than others and using these relative distance between events could help create a more informative disease progression model.

3. Discriminative event-based model

Fonteijn's and Huang's EBM are generative models where the likelihood $p(X|S)$ is maximized. Huang's EBM also estimates subject-specific ordering based on a generative approach. Here, we propose our novel method for estimating central ordering of events (S), a discriminative event-based model (DEBM).

The proposed framework is discriminative in nature, since we estimate s_j directly based on the posterior probabilities of individual biomarkers becoming abnormal. We also introduce a new concept of relative distance between events. This subsequently leads to a novel continuous patient staging algorithm. Fig. 2 shows the different steps involved in our approach.

In Section 3.1, we present the method to robustly estimate biomarker distributions in pre-event and post-event classes, given a single cross-sectional measurement of biomarkers. In Section 3.2, we present a way for estimating s_j, and we address the problem of estimating a disease timeline from noisy estimates of s_j. In Section 3.3, we present the continuous patient staging method.
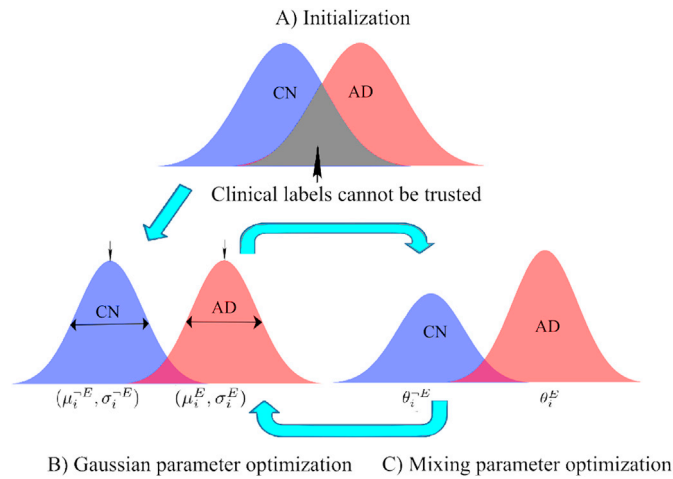
3.1. Biomarker progression

In this section, we propose a method to robustly convert $x_{j,i}$ to $p(E_i|x_{j,i})$, which denotes the posterior probability of a biomarker measurement being abnormal. Assuming a paradigm similar to that in previous EBM variants (Huang and Alexander, 2012; Young et al., 2014), the probability density functions (PDF) of pre-event ($p(x_{j,i}|\neg E_i)$) and post-event ($p(x_{j,i}|E_i)$) classes in the biomarkers are assumed to be represented by Gaussians, independently for each biomarker. There are two reasons why constructing these PDFs is non-trivial. Firstly, the labels (clinical diagnoses) for the subjects do not necessarily represent the true labels of all the biomarkers extracted from the subject. Not all biomarkers are abnormal for subjects with AD diagnosis, while some of the cognitively normal (CN) subjects could have undiagnosed pre-symptomatic conditions. Secondly, the clinical diagnosis can be non-binary and include classes such as MCI, with significant number of biomarkers in normal and abnormal classes.

In our approach we address these two issues independently. We make an initial estimate of the PDFs using biomarkers from easily classifiable CN and easily classifiable AD subjects and later refine the estimated PDF using the entire dataset.

A Bayesian classifier is trained for each biomarker using CN and AD subjects, based on the assumption that there are no biomarkers in the pre-symptomatic stage for CN subjects and all the biomarkers are abnormal for AD subjects. This classifier is subsequently applied to the training data, and the predicted labels are compared with the clinical labels. The misclassified data in the dataset could either be outliers in each class resulting from our aforementioned assumption or could genuinely belong to their respective classes and represent the tails of the true PDFs. Irrespective of the reason of misclassification, we remove them for initial estimation of the PDFs. This procedure thus results, for each biomarker, in a set of easily classifiable CN subjects (whose biomarker values represent normal values) and easily classifiable AD subjects (whose

---

[4] Fonteijn et al. (2012) briefly mention the idea of capturing relative distance between events, but it was not validated or used in any of the experiments.

A) Initialization

CN AD

Clinical labels cannot be trusted

CN AD CN AD

$(\mu_i^{-E}, \sigma_i^{-E})$ (μ_i^E, σ_i^E) θ_i^{-E} θ_i^E

B) Gaussian parameter optimization C) Mixing parameter optimization

Figure 3. Overview of the steps involved in the proposed Gaussian Mixture Model optimization strategy. A) Illustration of the initialization step for Gaussian Mixture Model. Rejecting the tails of the Gaussian distribution in CN and AD class is done to account for the fact that some of the CN subjects could be in pre-symptomatic stage of disease progression and some of the biomarkers could still be normal in AD subjects. B and C) This is followed by iterative estimation of Gaussian parameter optimization and Mixing parameter optimization.

biomarker values represent abnormal values). This is shown in the top part of Fig. 3.

As we use Gaussians to represent the PDFs, we calculate initial estimates for mean and standard deviation for both normal ($\mu_i^{-E}, \sigma_i^{-E}$) and abnormal classes ($\mu_i^E, \sigma_i^E$) based on 'easy' CN and 'easy' AD subjects for each biomarker i. As these means and standard deviations are estimated based on truncated Gaussians, these are biased estimates. The initial estimates of standard deviations are always smaller than the expected unbiased estimates whereas the initial estimates of means are under-estimated for Gaussians with smaller means (as compared to the other class for corresponding biomarkers) and overestimated for Gaussians with larger means.

We refine the initial estimates using a Gaussian mixture model (GMM) and include all the available data, including MCI subjects and previously misclassified cases. To obtain a robust GMM fit, a constrained optimization method is used, with bounds on the means, standard deviations and mixing parameters, based on the aforementioned relationship between the initial estimates and their corresponding expected unbiased estimates. The objective function for optimization for biomarker i is a summation of log-likelihoods, for all subjects:

$$C_i = \sum_{\forall j} \log f(x_{j,i}) \tag{5}$$

where the likelihood function $f(x_{j,i})$ is computed as a function of mixing parameters ($\theta_i^E, \theta_i^{-E}$) for the groups corresponding to post-event and pre-event respectively and their corresponding Gaussian distributions (μ_i^E, σ_i^E) and ($\mu_i^{-E}, \sigma_i^{-E}$):

$$f(x_{j,i}) = \theta_i^E p(x_{j,i}|\mu_i^E, \sigma_i^E) + \theta_i^{-E} p(x_{j,i}|\mu_i^{-E}, \sigma_i^{-E}) = \theta_i^E p(x_{j,i}|E_i) + \theta_i^{-E} p(x_{j,i}|\neg E_i) \tag{6}$$

θ_i^E and θ_i^{-E} are selected such that $\theta_i^E + \theta_i^{-E} = 1$. The mixing parameters and the Gaussian parameters are optimized alternately, until convergence of the mixing parameters. The initialization and optimization strategy in GMM is illustrated in Fig. 3.

The strategy of alternating between optimizing for mixing parameter and optimizing for Gaussian parameters in combination with the initialization strategy and the subsequent constraints is different from all

previous versions of EBM and it will be shown in Section 5 that this results in more accurate central ordering of events in most cases.

3.2. Estimating a disease progression timeline

3.2.1. Estimating subject-specific orderings

The PDF thus obtained is used for classification of the biomarkers using a Bayesian classifier, where the mixing parameters (θ_i^E and θ_i^{-E}) are used as the prior probabilities ($p(E_i)$ and $p(\neg E_i)$ respectively) when estimating posterior probabilities for each biomarker. We assume these posterior probabilities to be a measure of progression of a biomarker. Thus, s_j is established such that:

$$s_j \ni p\left(E_{s_j(1)}|x_{j,s_j(1)}\right) > p\left(E_{s_j(2)}|x_{j,s_j(2)}\right) > \dots > p\left(E_{s_j(N)}|x_{j,s_j(N)}\right) \tag{7}$$

Missing biomarker values are implicitly handled in this definition of s_j, as s_j only consists of events for which biomarkers are present for subject j. The posterior probabilities in Equation (7) are influenced not only by progression of the biomarker values to their abnormal states, but also by inherent variability in normal and abnormal biomarker values across subjects, and by measurement noise. Disentangling measurement noise and inherent variability in normal biomarker values from progression of the biomarker to its abnormal state can only be done based on longitudinal data. This makes s_j a noisy estimate.

3.2.2. Estimating a central ordering

Since the event ordering for each subject is estimated independently, any heterogeneity in disease progression is captured in the estimates of s_j. The central event ordering (S) is the mean of the subject-specific estimates of s_j. To describe the distribution of s_j, we make use of a generalized Mallows model. The generalized Mallows model is parameterized by a central ('mean') ordering as well as spread parameters (analogous to the standard deviation in a normal distribution). The central ordering is defined as the ordering that minimizes the sum of distances to all subject-wise orderings s_j. To measure distance between orderings, an often used measure is Kendall's Tau distance (Huang and Alexander, 2012). Kendall's Tau distance between a subject specific event ordering (s_j) and central ordering (S) can be defined as:

$$K(S, s_j) = \sum_{i=1}^{N-1} V_i(S, s_j) \tag{8}$$

where $V_i(S, s_j)$ is the number of adjacent swaps needed so that event at position i is the same in s_j and S. In case of missing biomarkers, $K(S, s_j)$ is computed for a subset of S consisting only of the events corresponding to the available biomarkers for subject j.

Since the estimates of s_j are based on rankings of posterior probabilities, it would be desirable to penalize certain swaps more than others, based on how close the posterior probabilities are to each other. To this end, we introduce a probabilistic Kendall's Tau distance, which penalizes each swap based on the difference in posterior probabilities of the corresponding events.

$$\widehat{K}(S, s_j) = \sum_{i=1}^{N-1} \widehat{V}_i(S, s_j) \tag{9}$$

$\widehat{V}_i \forall i \in [1, N-1]$ is computed sequentially using the following algorithm[5]:

Algorithm 1. Probabilistic Kendall Tau distance between Subject-specific event orderings and central event ordering

---

[5] The summation symbol in step 4 was missed accidentally in Venkatraghavan et al. (2017).

```
1:  for i ∈ [1, N - 1] do
2:      k ← s_j^{-1}(S(i))
3:      if k > i then
4:          V̂_i(S, s_j) ← ∑_{l=i+1}^{k} p_i - p_l
5:          Move s_j(k) to position i and update s_j
6:      else
7:          V̂_i(S, s_j) ← 0
```

Where p_a is shortened notation for $p(E_{s_j(a)}|x_{j,s_j(a)})$.

This variant of Kendall's Tau distance is quite close to the weighted Kendall's Tau distance defined in the permutation space introduced in Kumar and Vassilvitskii (2010). The difference stems from the fact that since the probabilistic Kendall's Tau distance is between individual estimates and a central-ordering, the penalization of each swap is weighted asymmetrically as $\widehat{V}_i(S, s_j) \neq \widehat{V}_i(s_j, S)$.

The optimum S is the one that minimizes $\sum_{\forall j} \widehat{K}(S, s_j)$. However, computing a global optimum S based on subject-wise orderings is NP-hard. Thus getting a good initial estimate of S is important to ensure the estimated S is not a suboptimal local optimum. In our implementation the initial estimate of S is based on ordering θ_i^{-E}. The motivation for this is discussed in Section 3.3. S was further optimized based on the algorithm introduced by Fligner and Verducci (1988) to estimate the central ordering.

3.2.3. Estimating event centers

The S that has been derived in this manner, is an estimate of the sequence in which the biomarkers become abnormal during the progression of a disease. However, it falls short of being a disease timeline, because it does not provide information about the proximity of consecutive events. To address this issue, we estimate distances between events by computing the cost of adjacent swaps in the event ordering, as measured by summation of probabilistic Kendall's Tau distance over all subjects.

$$\Gamma_{i+1,i} = \sum_{\forall j} \widehat{K}(S_{i+1,i}, s_j) - \widehat{K}(S, s_j) \tag{10}$$

where $S_{i+1,i}$ is identical to S except for the swap between events at locations i and $i + 1$, and $\Gamma_{i+1,i}$ is the cost of the swap. This represents the cost for the central ordering to be $S_{i+1,i}$ instead of S. We hypothesize that the closer the events $i + 1$ and i are to each other, the lower the swapping cost would be. Hence we consider these costs to be proportional to distance between events in terms of biomarker progression.

To estimate the distance of the first biomarker being abnormal (event) in S to a hypothetical disease-free individual, we introduce a pseudo-event which becomes abnormal at the beginning of the disease timeline and hence is abnormal for all the subjects in the database *i.e.* $p(E_0|x_{j,0}) = 1 \ \forall j$. Similarly, we introduce another pseudo-event which becomes abnormal at the end of the disease timeline and hence is normal for all the subjects in the database *i.e.* $p(E_{N+1}|x_{j,N+1}) = 0 \ \forall j$. We scale $\Gamma_{i+1,i} \forall i \in [0, N]$ such that $\sum \overline{\Gamma}_{i+1,i} = 1$. Event center ($\lambda_k$) of event k in S for $k > 0$, is computed as follows:

$$\lambda_k = \sum_{i=0}^{k-1} \overline{\Gamma}_{i+1,i} \tag{11}$$

In fact, the concept of event centers can also be extended to Fonteijn's EBM by computing the cost of adjacent swaps in the event ordering as the difference in log-likelihoods as follows:

$$\Gamma_{i+1,i} = \log(p(X|S)) - \log(p(X|S_{i+1,i})) \tag{12}$$

Extension of this concept to Huang's EBM is not straightforward and is beyond this paper's scope.

The set of event centers $\lambda_{1,2,...,N}$, will henceforth be referred to as Λ. This results in a disease timeline, with S giving information about the order of progression of biomarkers and Λ giving information about the event centers in this timeline.

3.3. Patient staging

Once the central ordering of events (S) and event centers (Λ) have been determined, we propose a patient staging algorithm where a patient stage (Υ_j) is interpreted as an expectation of λ_k with respect to the conditional distribution $p(k|S, X_j)$. Thus, Υ_j can be written as given below:

$$\Upsilon_j = \frac{\sum_{k=1}^{N} \lambda_k p(k|S, X_j)}{\sum_{k=1}^{N} p(k|S, X_j)} \tag{13}$$

Multiplying $p(S, X_j)$ in both numerator and denominator and using the chain rule of probability results in:

$$\Upsilon_j = \frac{\sum_{k=1}^{N} \lambda_k p(k, S, X_j)}{\sum_{k=1}^{N} p(k, S, X_j)} \tag{14}$$

Using chain rule of probability, we can write $p(k, S, X_j)$ as:

$$p(k, S, X_j) = p(X_j|k, S)p(k, S) \tag{15}$$

If we assume a uniform distribution of $p(k|S)$ and $p(S)$ as in Fonteijn et al. (2012), $p(k, S, X_j)$ becomes equal to $p(X_j|k, S)$, which was used for patient staging in Fonteijn's EBM as discussed in Section 2. However we use prior knowledge in order to define a more informative distribution $p(k, S)$:

$$p(k, S) = \frac{\prod_{i=1}^{k} \theta_{S(i)}^{E} \prod_{i=k+1}^{N} \theta_{S(i)}^{-E}}{Z} \tag{16}$$

where Z is a normalizing factor, chosen so as to make this a probability. This choice of $p(k, S)$ can be justified because biomarkers which become abnormal earlier in the disease process are more likely to have a higher value of θ_i^E than the biomarkers which become abnormal later. Hence it is far more likely to have a central-ordering based on ascending values of θ_i^{-E} than an ordering with ascending values of θ_i^E. It should be noted that, the choice of $p(k, S)$ is not unique. For example, it could also be any *n*-th power of the above equation $\forall n > 0$. Thus, from Equation (15), 16 and 4, we get:

$$p(k, S, X_j) \propto \prod_{i=1}^{k} p(x_{j,S(i)}|E_{S(i)}) \theta_{S(i)}^{E} \times \prod_{i=k+1}^{N} p(x_{j,S(i)}|\neg E_{S(i)}) \theta_{S(i)}^{-E} \tag{17}$$

Using the above value of $p(k, S, X_j)$ in Equation (14), results in continuous patient stages.

4. Experiments

This section describes the experiments performed to benchmark the accuracy of the proposed DEBM algorithm and compare it with state-of-the-art EBM methods. The EBM methods used for comparison in these experiments are Huang's EBM (Huang and Alexander, 2012) and the variant of Fonteijn's EBM that is suited for AD disease progression modeling (Young et al., 2014). The source code for DEBM and Fonteijn's EBM, with different mixture modeling techniques and patient staging techniques discussed in this paper have been made publicly available online under the GPL 3.0 license: https://github.com/88vikram/pyebm/ . The source code for Huang's EBM used in our experiments was provided by the authors of the method.

For brevity, Fonteijn's EBM and Huang's EBM will henceforth be referred to as FEBM and HEBM, respectively. The mixture model used with an EBM model (as the one described in Section 3.1) will be denoted by a subscript. For example, FEBM with the Gaussian mixture model proposed in Young et al. (2014) will be referred to as FEBM_May. The Gaussian mixture model optimization techniques in Huang and Alexander (2012), Venkatraghavan et al. (2017) and the one introduced in

Table 1
Demographics for the whole population.

Demographics				
Diagnosis	n	Sex M/F	Age [yrs.]	Edu. [yrs.]
CN	417	209/208	74.76 ± 5.72	16.28 ± 2.73
SMC	106	44/62	72.20 ± 5.53	16.76 ± 2.51
MCI	872	515/357	73.00 ± 7.61	15.90 ± 2.83
AD	342	189/153	75.02 ± 7.78	15.17 ± 2.98

Table 2
Demographics for the homogeneous subset of subjects.

Demographics				
Diagnosis	n	Sex M/F	Age [yrs.]	Edu. [yrs.]
CN	160	83/77	73.56 ± 5.81	16.38 ± 2.66
MCI	414	249/165	73.20 ± 7.11	16.01 ± 2.79
AD	216	125/91	74.36 ± 8.06	15.45 ± 2.94

Table 3
Abbreviations used in Figs. 5 and 6 along with their full names (Hammers et al., 2003).

Abbreviation	Full name
L	Left
R	Right
PHA	Parahippocampalis et Ambiens
Med.	Medial
Inf.	Inferior
Sup.	Superior
Temp.	Temporal
Pos.	Posterior
Lat.	Lateral
Ant.	Anterior
OT	Occipitotemporal
Cent.	Central
Mid.	Middle
Rem.	Remainder
Occ.	Occipital
PS	Pre-subgenual

this paper will be denoted with subscripts 'jh', 'vv1' and 'vv2' respectively.[6]

Data used in the experiments were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu)[7]. We begin with the details of the experiments performed on ADNI data to estimate the event ordering in Section 4.1. Since the ground-truth event ordering is unknown for clinical datasets, we resort to using the ability of patient staging to classify AD and CN subjects, as an indirect way of measuring the reliability of the event ordering. We also measure the accuracy of event ordering and relative distance between events more directly by performing extensive experiments on synthetic data simulating the progression of AD. The details of these experiments are given in Section 4.2.

4.1. ADNI data

We considered 1737 subjects from ADNI 1, Go and 2 (417 CN, 106 with Significant Memory Concern (SMC), 872 MCI and 342 AD subjects) who had a structural MRI (T1w) scan at baseline. Study subject demographics are summarized in Table 1. The T1w scans were non-uniformity corrected using the N3 algorithm (Tustison et al., 2010). This was followed by multi-atlas brain extraction using the method described in Bron et al. (2014). Multi-atlas segmentation was performed (Hammers et al., 2003; Gousias et al., 2008) using the structural MRI scans to obtain a region-labeling for 83 brain regions in each subject using a set of 30 atlases. Probabilistic tissue segmentations were obtained for white matter, gray matter (GM), and cerebrospinal fluid on the T1w image using the unified tissue segmentation method (Ashburner and Friston, 2005) of SPM8 (Statistical Parametric Mapping, London, UK). The probabilistic GM segmentation was then combined with region labeling to obtain GM volumes in the extracted regions. We also downloaded CSF (A β_{1-42} (ABETA), TAU and p-TAU) and cognitive score (MMSE, ADAS-Cog) values from the ADNI database, making the total number of features equal to 88.

The features TAU and p-TAU were transformed to logarithmic scales to make the distributions less skewed. GM volumes of segmented regions were regressed with age, sex and intra-cranial volume (ICV) and the effects of these factors were subsequently corrected for, before being used as biomarkers. The effect of age and sex was regressed out of CSF based features, whereas effects of age, sex and education was regressed out of cognitive scores.

We retained 52 biomarkers (GM volume based biomarkers of 47 regions, 3 CSF and 2 cognitive scores) having significant differences between CN and AD subjects using Student's t-test with $p < 0.005$, after Bonferroni correction. These biomarker values were used to perform Experiments 1 and 2.

Experiment 1(a): A subset of 7 biomarkers including the 3 CSF features, MMSE score, ADAS-Cog score, gray matter volume of the hippocampus (combined volume of left and right hippocampi) and gray matter volume in whole brain was created. Event ordering of these 7 biomarkers was inferred using DEBM. We studied the positional variance of central ordering and variance of event centers inferred by DEBM by creating 100 bootstrapped samples of the data.

Experiment 1(b): The Biomarkers were ranked based on their aforementioned p-value and the above experiment was repeated with top 25 and top 50 biomarkers to investigate if the event-centers estimated for the subset of Biomarkers used in Experiment 1(a), remain comparable to the ones estimated in Experiment 1(a).

Experiment 2: As an indirect way of measuring the accuracy of the estimated event ordering, we use patient staging based on the estimated event orderings as a way to classify CN and AD subjects in the database. 10-fold cross validation was used for this purpose. AUC measures were used to measure the performance of these classifications and thus indirectly hint at the reliability of the event ordering based on which the corresponding patient staging were performed.

We used varying number of biomarkers (ranked based on their p-value) ranging from 5 to 50 in steps of 5 for this experiment. We used the methods FEBM$_{ay}$, HEBM$_{jh}$, DEBM$_{vv1}$ and DEBM$_{vv2}$ for inferring the ordering. Patient staging was done based on the methods described in their respective papers. Since the earlier version of DEBM (Venkatraghavan et al., 2017) had not introduced a patient staging method, we use the patient staging method described in this paper for evaluating the method.

Experiment 3(a): To study disease progression in a homogeneous population showing signs of typical AD progression, Experiment 1(a) was repeated with a subset of subjects, selected based on their CSF ABETA values. For this experiment, we selected ABETA positive MCI and AD subjects (ABETA < 192 pg/ml) and ABETA negative CN subjects (ABETA

---

[6] Mixture model 'ay' optimizes for Gaussian and mixing parameters together. Initialization of Gaussian parameters for optimization is done without rejecting the overlapping part of Gaussians in CN and AD classes. 'vv1' also optimizes for Gaussian and mixing parameters together (although with much stricter bounds) but the initialization of Gaussian parameters is similar to the one in this paper. 'jh' couples mixture modeling with estimation of subject-specific ordering to estimate a combined optimum solution.

[7] 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 Alzheimers disease (AD). For up-to-date information, see www.adni-info.org.

$>= 192$ pg/ml). This cut-off was chosen according to the results of Shaw et al. (2009). Moreover, we excluded all SMC subjects and subjects with missing ABETA biomarker values. This subset of subjects will henceforth be referred to as the 'homogeneous subset'. Demographics for the homogeneous subset are summarized in Table 2. We excluded ABETA biomarker when inferring the event ordering using DEBM.

Experiment 3(b): We retained 49 biomarkers (GM volume based biomarkers of 45 regions, 2 CSF biomarkers excluding ABETA and 2 cognitive scores) having significant differences between CN and AD subjects in the homogeneous subset using Student's t-test with $p < 0.05$, after Bonferroni correction. The biomarkers were ranked based on their aforementioned p-value and the above experiment was repeated with top 24 and top 49 biomarkers, to investigate if the event-centers estimated for the subset of biomarkers used in Experiment 3(a), remain comparable to the ones estimated in Experiment 3(a).

4.2. Simulation data

We used the framework developed by Young et al. (2015a) for simulating cross-sectional data consisting of scalar biomarker values for CN, MCI and AD subjects. In this framework, disease progression in a subject is modeled by a cascade of biomarkers becoming abnormal and individual biomarker trajectories are represented by a sigmoid. The equation for generating biomarker values for different subjects is given below:

$$x_{j,i}(\Psi) = \frac{R_i}{1 + \exp\left(-\rho_i\left(\Psi - \xi_{j,i}\right)\right)} + \beta_{j,i} \tag{18}$$

Ψ denotes disease stage of a subject which we take to be a random variable distributed uniformly throughout the disease timeline. ρ_i signifies the rate of progression of a biomarker, which we take to be equal for all subjects. $\xi_{j,i}$ denotes the disease stage at which the biomarker becomes abnormal. $\beta_{j,i}$ denotes the value of the biomarker when the subject is normal and R_i denotes the range of the sigmoidal trajectory of the biomarker, which we take to be equal for all subjects.

In our experiments, $\beta_{j,i}$ and $\xi_{j,i}$ $\forall j$ are assumed to be random variables with Normal distribution $\mathbb{N}(\mu_{\beta_i}, \Sigma_{\beta_i})$ and $\mathbb{N}(\mu_{\xi_i}, \Sigma_{\xi_i})$ respectively. μ_{β_i} is equal to the mean value of the corresponding biomarker in the CN group of the selected ADNI data. R_i is equal to the difference between the mean values of the biomarker in the CN and AD groups of the selected ADNI data. Σ_{β_i} represents the variability of biomarker values in the CN group. We consider a relative scale for Σ_{β_i}, where 1 refers to the observed variation among the CN subjects in ADNI data. Variation in $\xi_{j,i}$ is controlled by Σ_{ξ_i} and results in variation in ordering among subjects in population and could be seen as a parameter controlling the disease heterogeneity within a simulated population. Σ_{ξ_i} $\forall i$ is varied in multiples of $\Delta\xi$, where $\Delta\xi$ is the average difference between adjacent μ_{ξ_i}. μ_{ξ_i} refers to the event centers of various biomarkers. The set of μ_{ξ_i} $\forall i$ will collectively be referred to as Λ_{gt} and they will be used to assess the accuracy of estimated event centers (λ_i).

The parameters in the simulation framework that could have an effect on the performance of EBMs are Σ_{β_i}, μ_{ξ_i}, Σ_{ξ_i}, and ρ_i. Apart from this, the number of subjects (M) and the number of biomarkers (N) in the dataset could also have an effect on the performance of EBMs. Using this simulation framework, we study the effect of the aforementioned parameters on the ability of different variants of EBM algorithms to accurately infer the ground-truth central ordering in the population. Change in μ_{β_i} results only in a translational effect on biomarker values and change in R_i results only in a scaling effect on biomarker values. These factors do not affect the performance of the EBMs and hence were not evaluated in our experiments.

Performance of an EBM method can be measured using error in estimation of either S or Λ. Error in estimating S (ε_S) will henceforth be referred to as 'ordering error' whereas the error in estimating Λ (ε_Λ) will

henceforth be referred to as 'event-center error'. ε_S is computed using the following equation:

$$\varepsilon_S = \frac{K\left(S, S_{gt}\right)}{\binom{N}{2}} \tag{19}$$

where S_{gt} is the ground truth ordering. ε_S is effectively a normalized Kendall's Tau distance between S and S_{gt}. The normalization factor for $\binom{N}{2}$, was chosen to make the accuracy measure interpretable for different number of biomarkers.

For comparing Λ and Λ_{gt}, Λ were scaled and translated such that the mean and standard deviation of Λ were equal to that of Λ_{gt}. This is done because we are only interested in evaluating the errors in estimating relative distance between events and not the absolute position of event-centers. The choice of scale in event-centers are arbitrary and the chosen scale for the estimated event-centers was based on pseudo-events, which need not necessarily coincide with the simulation framework's ground-truth event-centers.

$$\varepsilon_\Lambda = \sum_{\forall i}\left|\lambda_i^{st} - \mu_{\xi_i}\right| \tag{20}$$

where λ_i^{st} is the scaled and translated version of λ_i.

As mentioned before, the factors that can have an effect on the performance of EBMs are Σ_{β_i}, μ_{ξ_i}, Σ_{ξ_i}, ρ_i, M and N. In each of the following 5 experiments, a few of these factors were varied while the others were set to their default values. The default value for Σ_{β_i} was taken to be 1 as this corresponds to the observed variation among CN subjects in ADNI. μ_{ξ_i} were spaced equidistantly, i.e., $\mu_{\xi_{i+1}} - \mu_{\xi_i} = 1/(N+1)$. As the actual variation in event centers among different subjects is not known in a clinical dataset, the default value of Σ_{ξ_i} was taken to be $2\Delta\xi$. For the sake of simplicity of notation $\Delta\xi$ will be omitted henceforth, and the values of Σ_{ξ_i} are implicitly in multiples of $\Delta\xi$. ρ_i was considered to be equal for all biomarkers by default. The default values for M and N were 1737 and 7 respectively, mimicking the dataset used in Experiment 1(a). For each simulation setting, 50 repetitions of simulation data were created and used for benchmarking the performance of EBMs on synthetic data.

Experiment 4: The first simulation experiment was performed to study the effect of $\Sigma_\beta \in [0.2, 1.8]$ and $\Sigma_\xi \in [0, 4]$, varying one at a time while keeping the other at its mean value. The ε_S of FEBM$_{ay}$, FEBM$_{vv2}$, HEBM$_{jh}$, HEBM$_{vv2}$, DEBM$_{vv1}$ and DEBM$_{vv2}$ were determined.

Experiment 5: The above experiment was repeated for DEBM$_{vv2}$ and FEBM$_{vv2}$ and the ε_Λ were measured for the two methods.

Experiment 6: This experiment was performed to study the effect of a non-uniform distribution of μ_{ξ_i}. Σ_β and Σ_ξ combinations of $(0.6, 1)$, $(1.0, 2)$, $(1.4, 3)$ and $(1.8, 4)$ were tested to study their effect in non-uniformly spaced biomarkers. ε_S of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ were measured. Additionally, ε_Λ of DEBM$_{vv2}$ and FEBM$_{vv2}$ were measured. To also study the effect of unequal rates of progression of biomarkers (ρ_i), the above experiment was performed once with equal ρ_i for all biomarkers and once when they were unequal. The experiment with unequal biomarker rates had the same mean biomarker progression rate as the experiment with equal biomarker rates. The progression rates of different biomarkers has been included as supplementary material (Fig. S1).

Experiment 7: This experiment was performed to study the influence of the number of subjects (M). M was varied from 100 to 2100 in steps of 200. ε_S of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ were measured. DEBM$_{vv2}$ and FEBM$_{vv2}$ were also assessed based on ε_Λ.

Experiment 8: This experiment was performed to study the influence of the number of biomarkers (N). N was varied from 7 to 52 in steps of 5. In each random generation of a dataset, we randomly selected (with replacement) the biomarkers to be used in the iteration. This was done to study the effect of N on the EBM models and separate it from the effect of
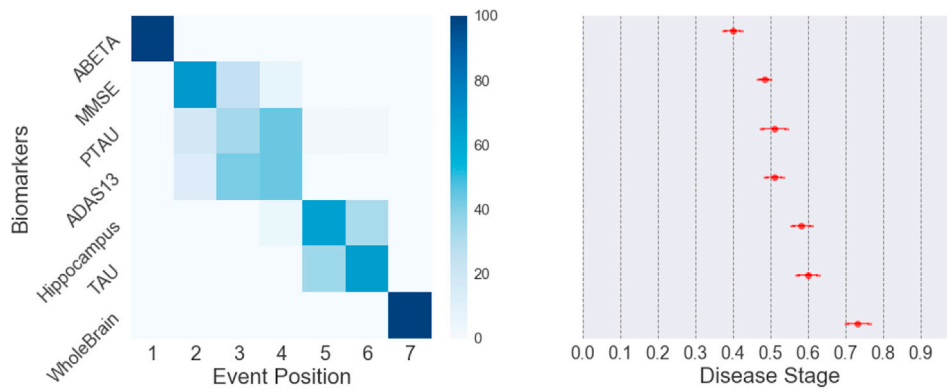
Fig. 4. Experiment 1(a): DEBM$_{vv2}$ with 7 Events. The positional variance diagram (left) shows the uncertainty in estimating the central event ordering. The event-center variance diagram (right) shows the standard error of estimated event centers. These were measured by 100 repetitions of bootstrapping.
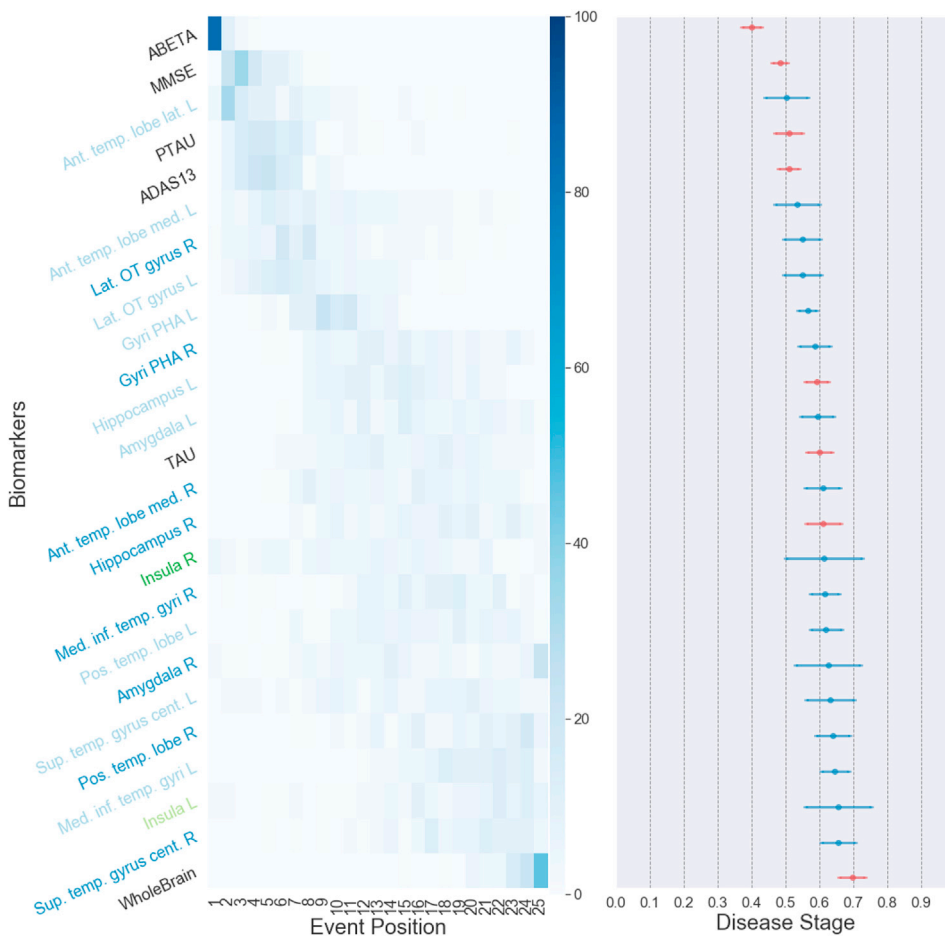
Fig. 5. Experiment 1(b): DEBM$_{vv2}$ with 25 Events. The positional variance diagram (left) shows the uncertainty in estimating the central event ordering and the event-center variance diagram (right) shows the standard error of estimated event centers. These were measured by 100 repetitions of bootstrapping. The event centers of the biomarkers used in Fig. 4 are marked in red. Table 3 shows the full forms of the abbreviations used in the y-axis labels. Fig. 7 maps the colors used for y-axis labels to different lobes in the brain.

adding weaker biomarkers. ε_S of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ were measured. DEBM$_{vv2}$ and FEBM$_{vv2}$ were also assessed based on ε_Λ.

5. Results

5.1. ADNI data

Experiment 1: Fig. 4 shows the positional variance and event-center variance obtained using DEBM$_{vv2}$ with 7 events. The Gaussian mixture model parameters for every biomarker have been tabulated in the supplementary material, Table S1.

It can be seen from Fig. 4 (left) that CSF-based biomarkers ABETA

becomes abnormal before MMSE and CSF-based p-TAU. This is followed by ADAS13, Hippocampal volume, TAU and whole brain volume events. However Fig. 4 (right) shows that the event centers for MMSE, ADAS13, p-TAU are close to each other and so are the event-centers of TAU and hippocampus volume. The event associated with the TAU biomarker seems closer to the whole brain volume event as they are in positions 6 and 7 of Fig. 4 (left). However, the centers of these two events are quite far apart in Fig. 4 (right) and the p-TAU event (position 2) is closer to the TAU event than whole brain volume event.

As the number of biomarkers increases, the variation in the positions also increases considerably, as seen in Fig. 5 (left) and 6 (left). The event centers of the biomarkers used in Experiment 1(a) remain fairly

Fig. 6. Experiment 1(b): DEBM$_{vv2}$ with 50 Events. Positional variance diagram (left) shows the uncertainty in estimating the central event ordering and event center variance diagram (right) shows the standard error of estimated event-centers. These were measured by 100 repetitions of bootstrapping. The event-centers of the biomarkers used in Fig. 4 are marked in red, whereas the ones used in Fig. 5 are marked in blue. Table 3 shows the full forms of the abbreviations used in the y-axis labels. Fig. 7 maps the colors used for y-axis labels to different lobes in the brain.

consistent (± 0.05) in Experiment 1(b). It can also be seen that bio-markers with lower *p*-values (biomarkers included in the model with 50 biomarkers and not in the model with 25 biomarkers), have larger variance in their event-center estimation.

Experiment 2: Fig. 8 (a) shows the mean AUC when using patient stages for classifying CN versus AD subjects using DEBM and other variants of EBM methods. It can be observed that the AUC of all the methods decreases as the number of events increases. The proposed method DEBM$_{vv2}$ followed by the proposed patient staging algorithm outperforms all the existing EBM variants consistently.

Fig. 7. Legend for the colors used in Fig. 5 and 6. The colors map different biomarker labels to lobes in the brain.
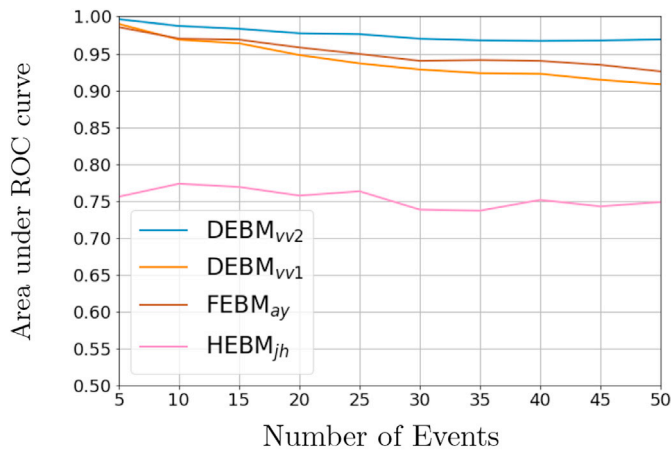
(a)
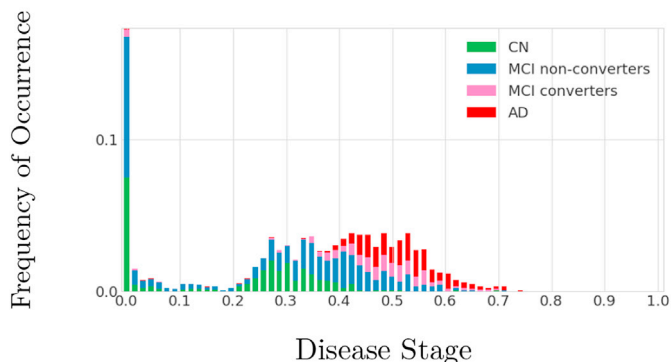
(b)

Fig. 8. Experiment 2: In (a) we see the variation of AUC with respect the number of biomarkers used for building the model using DEBM, when the obtained patient stages were used for classification of CN versus AD subjects. The AUC measure was obtained using 10-fold cross-validation. In (b) we see the frequency of occurrence of subjects in different disease stages, when the most significant 25 features were given as input to DEBM$_{vv2}$ for inferring the ordering as well as for patient staging.

Fig. 8 (b) shows the distribution of patient stages for the whole population when the most significant 25 features were given as input to DEBM$_{vv2}$. This graph shows a peak at disease stage 0 dominated by CN and MCI non-converters, which shows that these subjects are not progressing towards AD. The non-zero lower disease stages are dominated by CN subjects and MCI non-converters, whereas MCI converters[8] and the subjects with AD have higher disease stages.

Experiment 3: Fig. 9 shows the positional variance and event-center

---

[8] MCI converters are subjects who convert to AD within 3 years of baseline measurement.

variance obtained using DEBM$_{vv2}$ with 6 events, in the homogeneous subset of subjects. It can be seen from Fig. 9 that in the homogeneous subset of subjects, p-TAU event occurs before ADAS13 and MMSE events as opposed to p-TAU event occurring after ADAS13 and MMSE in Fig. 4. It can also be seen from Fig. 9 that the TAU event precedes Hippocampus volume event as opposed to Hippocampus event preceding the TAU event in Fig. 4.

The results of Experiment 3(b) with 24 and 49 have been included as supplementary material (Figs. S2 and S3).

5.2. Simulation data

Experiment 4: Fig. 10 shows the ordering errors of DEBM, FEBM and HEBM models with different mixture models as Σ_β and Σ_ξ increase. The error-bars depict mean and standard deviation of the errors obtained in 50 repetitions of simulations. It can be seen that the proposed optimization technique improves the performance of all three EBM models. The change is particularly evident when comparing the performance of FEBM$_{vv2}$ and FEBMay.

It can also be seen that FEBM$_{vv2}$ performs slightly better than DEBM$_{vv2}$ when Σ_ξ is low, but as Σ_ξ increases, the performance of FEBM$_{vv2}$ degrades significantly. The performance of HEBM is almost always worse than its FEBM or DEBM counterpart.

Experiment 5: Fig. 11 (a) and (b) shows the event-center errors in DEBM$_{vv2}$ and FEBM$_{vv2}$ as the variability in population (Σ_β) and disease heterogeneity (Σ_ξ) increases respectively. It should be noted from Figs. 10(b) and Figure 11 (b) that, even when the FEBM$_{vv2}$ gets the ordering more accurately than DEBM$_{vv2}$ in cases of low Σ_ξ, the event-center estimation of DEBM$_{vv2}$ is on par with or better than its FEBM counterpart.

Fig. 11 (c) shows the estimated event-center locations for $\Sigma_\beta = 1.0$ and $\Sigma_\xi = 2$ and the ground truth event-centers.

Experiment 6: Fig. 12 (a) shows the ordering errors of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ as Σ_β and Σ_ξ increase, when the ground-truth event centers (μ_{ξ_i}) are non-uniformly spaced. The spacing of μ_{ξ_i} can be observed in Fig. 12 (b), where the ground truth event-centers as well as the estimated event-centers of DEBM$_{vv2}$ and FEBM$_{vv2}$ are shown for $\Sigma_\beta = 1.0$ and $\Sigma_\xi = 2$. It can be observed that the estimated event-centers for DEBM$_{vv2}$ are much closer to the ground-truth event centers than those of FEBM$_{vv2}$ and also have a much lower variance over different iterations of simulations.

Fig. 12 (c) shows the ordering errors as Σ_β and Σ_ξ increases, when μ_{ξ_i} is non-uniformly spaced and ρ_i is not identical for all biomarkers. It should also be noted that the mean of ρ_i over all i has not changed between (a) and (c). The variation of errors in (c) is quite similar to the one in (a). This shows that performance of EBM methods that are reported in other experiments (where ρ_i is equal for all biomarkers) can be expected to not deteriorate in the more realistic scenario of ρ_i not being equal for all biomarkers. The event-center variance for $\Sigma_\beta = 1.0$ and $\Sigma_\xi = 2$ for the case of unequal ρ_i is very similar to (b) and has been included as supplementary material (Fig. S4).

Experiment 7: Fig. 13 shows the mean ordering errors of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ as a function of number of subjects in the dataset on one vertical axis and shows the mean event-center errors of DEBM$_{vv2}$ and FEBM$_{vv2}$ on the other vertical axis. As expected, the models perform better as the number of subjects increases. DEBM$_{vv2}$ is slightly better at inferring the central ordering than FEBM$_{vv2}$ when the number of subjects is very low, but FEBM$_{vv2}$ outperforms DEBM$_{vv2}$ when the number of subjects is higher. However, when the accuracy of event centers are considered, DEBM$_{vv2}$ consistently outperforms FEBM$_{vv2}$.

Experiment 8: Fig. 14 shows the mean ordering errors of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ as a function of the number of events (biomarkers) in the dataset on one vertical axis and shows the mean event-center errors of DEBM$_{vv2}$ and FEBM$_{vv2}$ on the other vertical axis. The biomarkers were selected randomly after replacement so that the chances
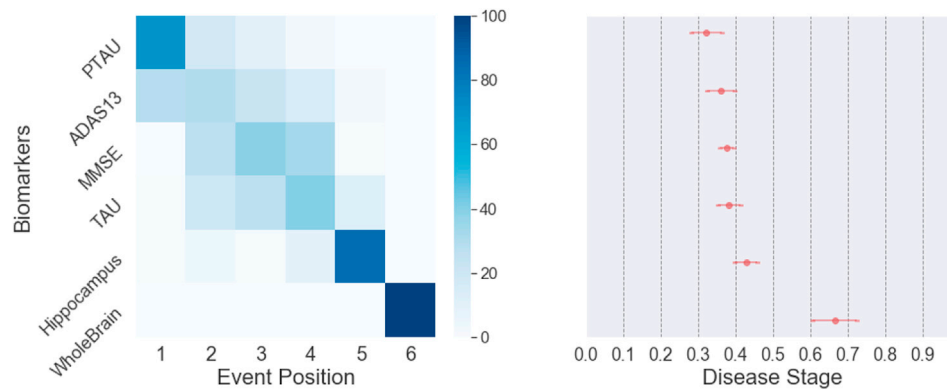
Fig. 9. Experiment 3(a): DEBM$_{vv2}$ with 6 Events, in the homogeneous subset of subjects. The positional variance diagram (left) shows the uncertainty in estimating the central event ordering. The event-center variance diagram (right) shows the standard error of estimated event centers. These were measured by 100 repetitions of bootstrapping.

of selecting a bad biomarker remain equal as the number of events increases. It can be noted that the errors of the EBM models increase as the number of events increases initially, even when the average quality of biomarkers remains the same. However the errors stabilize beyond a certain point and do not increase any more.

6. Discussion

We proposed a novel discriminative EBM framework to estimate the ordering in which biomarkers become abnormal during disease progression, based on a cross-sectional dataset. The proposed framework outperforms state-of-the-art EBM techniques in estimating the event ordering. We also introduced the concept of relative distance between event-centers, which enables creating a disease progression timeline. This in turn led to the development of a new continuous patient staging mechanism. In addition to the framework, we also proposed a novel probabilistic Kendall's Tau distance metric and a robust biomarker distribution estimation algorithm. In this section, we discuss different aspects of the proposed algorithm.

6.1. Event centers

Event-centers capture relative distance between events. This helps in creating the disease progression timeline from an ordering of events. If an event (Event A) leads to another event (Event B), this would be observed as event-center for A occurring before event-center for B. However EBMs cannot assess causality, and cannot distinguish the aforementioned case from the case when Event B is caused by some external factor which happened to occur after Event A.

Event centers are an intrinsic property of the biomarker used, for the selected population. This was observed in Experiment 1(b) where the event-centers estimated using DEBM$_{vv2}$ remained fairly consistent (± 0.05) across models using different number of biomarkers.

The estimated disease progression timeline can be used for inferring progression of the disease, with the event centers being synonymous to milestones of progression. A strict quantization of position in ordering of events (as reported in Oxtoby and Alexander (2017); Venkatraghavan et al. (2017); Young et al. (2015a); Young et al. (2014); Fonteijn et al. (2012)) in the positional variance diagram can sometimes be non-intuitive in terms of inferring actual progression of the disease. This was seen in Experiment 1, where the event center variance diagram showed that the TAU event (at position 6) was closer to the p-TAU event (at position 2) than the whole brain event (position 7).

The approach of scaling the event-centers between $[0, 1]$ has its advantages and disadvantages. The advantage of such a scaling is that models built on different biomarkers, but within the same population, remain comparable. For example, a model built with CSF and MRI based

biomarkers can be compared with a model built on MRI based biomarkers alone, as the event-centers of MRI based biomarkers would approximately be the same. On the other hand, the position of the first event relies heavily on the number of 'true' controls in the dataset (CN subjects who are not in an early asymptomatic stage of the disease). This is the result of introducing pseudo-events for scaling the events-centers.

Comparison of the event centers across different datasets with different number of controls (albeit with the same biomarkers) can be done in three ways. Event-centers can be scaled and translated such that the mean and standard deviation of event centers computed across different datasets are the same (similar to the comparisons between estimated and ground-truth event centers in this paper). Alternately, the event center of the first biomarker can be set as 0 and the event center of the last biomarker can be set to 1, before comparison. Lastly, in a dataset where controls (i.e., subjects whose biomarker values are all normal) can be easily identified, it would be better to exclude them for event-center computation.

The estimated event centers have a good correlation with the groundtruth disease timeline. This can be seen in the simulation experiments with and without uniform spacing of events (Experiments 5 and 6). It must however be noted that, the disease stages Ψ of the simulated subjects were distributed uniformly throughout the disease timeline. If the distribution is not uniform, we expect it to have an effect on the estimation of event centers. Analyzing the exact effect of such non-uniform distributions on the estimation of event centers and ways to estimate event centers invariant to the distribution of subjects on the disease timeline could be an interesting extension of the current work.

Experiment 6 also showed that different biomarkers having different rates of progression does not degrade the performance of EBM models, as long as the mean rate of progression is the same. We did not perform an experiment to benchmark the accuracies by changing the mean rates of progression of biomarkers. This experiment was already performed in Young et al. (2015b) and it was observed that FEBM ordering error decreases as the mean rates of progression increase.

FEBM assumes that the disease is homogeneous, as it expects all the subjects in the dataset to follow the same ordering. When the variability of ordering in different subjects is low, FEBM with the proposed mixture model 'vv2' outperforms DEBM with the proposed mixture model. This can be seen in the results of Experiments 4, 6 and 7. When the assumption becomes too restrictive, DEBM with the proposed mixture model outperforms FEBM. Even when the assumption holds true, estimation of event-centers with DEBM is more accurate than with FEBM.
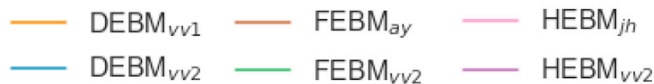
6.2. Patient staging

Existing patient staging algorithms discretize the patient stages based on event position, whereas the patient staging algorithm introduced in

(a)

(b)

DEBM$_{vv1}$ FEBM$_{ay}$ HEBM$_{jh}$

DEBM$_{vv2}$ FEBM$_{vv2}$ HEBM$_{vv2}$
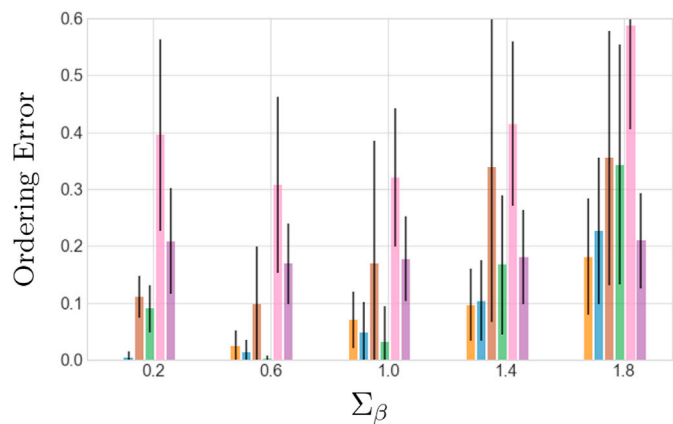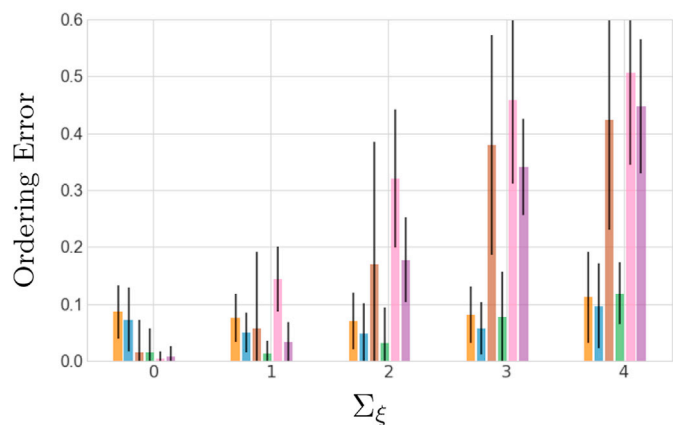
(c)

Fig. 10. Experiment 4: Ordering errors of DEBM$_{vv1}$, DEBM$_{vv2}$, FEBMay, FEBM$_{vv2}$, HEBMjh and HEBM$_{vv2}$ for 50 repetitions of simulations. Figure (a) shows the ordering error as a function of variability in population (Σ_β). Figure (a) shows the ordering error as a function of variation in ordering (Σ_ξ). Error bars in (a) and (b) represent standard deviations over the 50 repetitions. Figure (c) shows the legend for the plots in (a) and (b).

this paper takes relative distance between events into consideration while staging new subjects. This makes patient stages more useful for diagnosis and prognosis as they correlate more with the actual disease progression timeline. Discrete patient stages without considering the event centers could diminish the prognosis value of the obtained stages.

The cross-validation experiment on ADNI data (Experiment 2) showed that the CN and AD subjects are well separated after patient staging and that the AUC of the proposed method is better than that of the state-of-the-art EBM techniques. It also showed that MCI converters and non-converters are well separated after patient staging, without explicitly training the model to achieve this.

It must however be noted that even though heterogeneity of the disease was considered while inferring the central ordering, it was not considered for patient staging. Inferring multiple central orderings corresponding to different disease subtypes (Young et al., 2015a) and
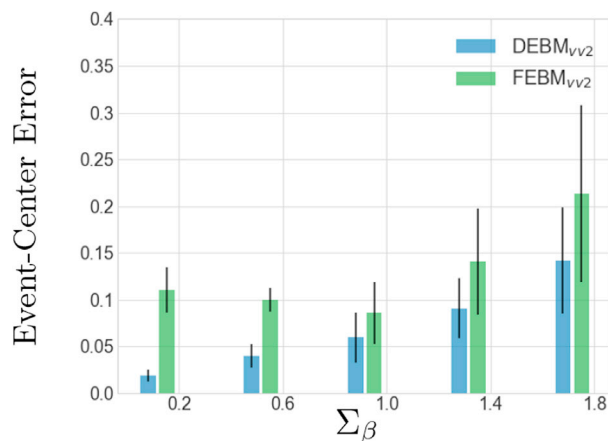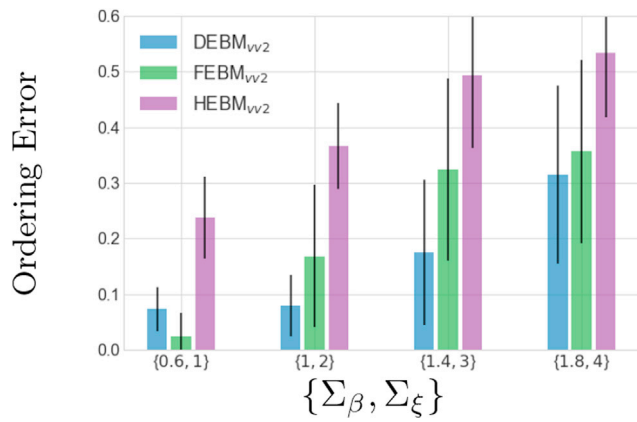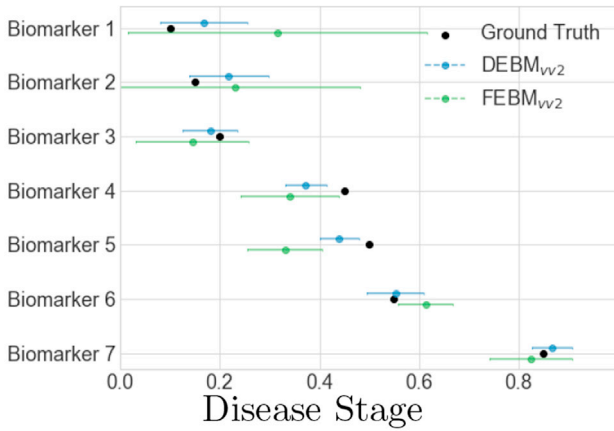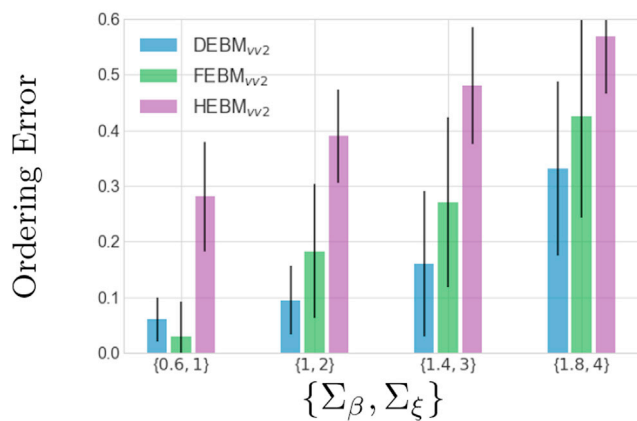

(a)
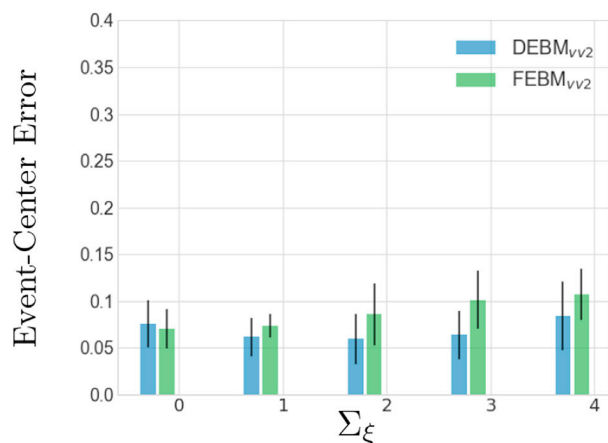
(b)
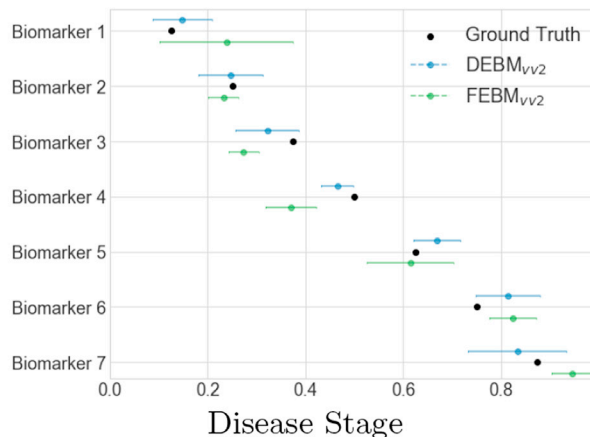
(c)

Fig. 11. Experiment 5: Figures (a) and (b) show the event-center errors of DEBM$_{vv2}$ and FEBM$_{vv2}$ as a function of Σ_β and Σ_ξ respectively. Figure (c) shows the estimated event-center locations for both methods as well as the ground-truth event centers. Error bars in (a), (b) and (c) represent standard deviation over 50 repetitions of simulation.

(a)

(b)

(c)

Fig. 12. Experiment 6: Figures (a) and (c) show the ordering errors of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ when μ_{ξ_i} are not uniformly distributed. Σ_β and Σ_ξ increase as we move from left to right. Figure (a) shows the errors in the case when ρ_i are identical for all the biomarkers whereas (c) shows the errors when ρ_i are different. Figure (b) shows the non-uniform μ_{ξ_i} as well as the estimated event-centers by DEBM$_{vv2}$ and FEBM$_{vv2}$ for the case of ρ_i being equal. Error bars in (a), (b) and (c) represent standard deviation over 50 repetitions of simulation.
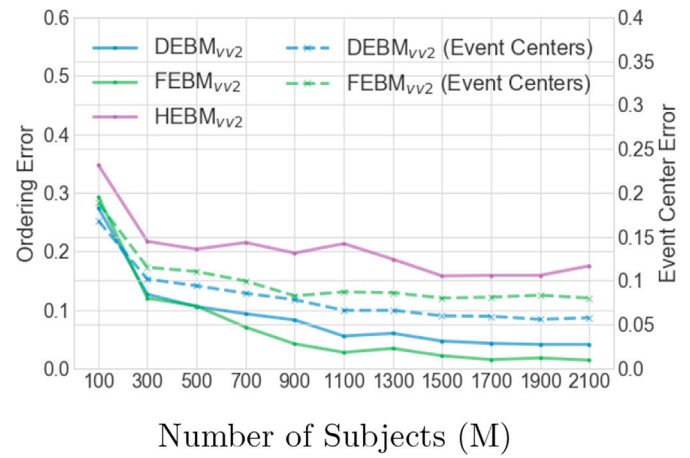
Fig. 13. Experiment 7: Ordering errors of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ as a function of number of subjects (M) in the dataset. It also shows the event-center errors of DEBM$_{vv2}$ and FEBM$_{vv2}$ as a function of M.
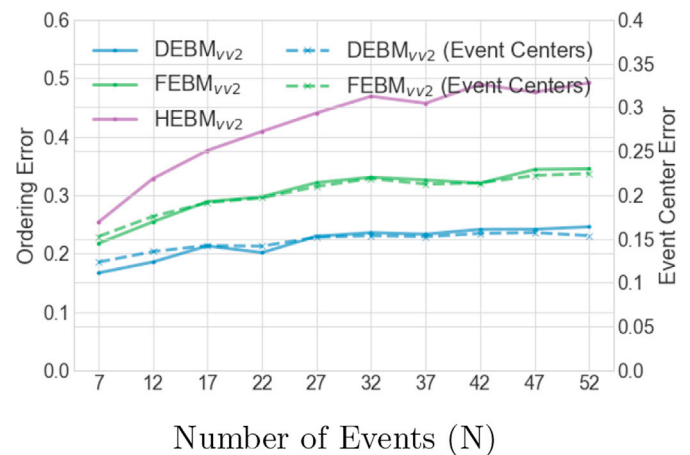
Fig. 14. Experiment 8: Ordering errors of DEBM$_{vv2}$, FEBM$_{vv2}$ and HEBM$_{vv2}$ as a function of number of events (N) in the dataset. It also shows the event-center errors of DEBM$_{vv2}$ and FEBM$_{vv2}$ as a function of N.

staging patients on one of these central orderings may help us overcome this drawback. Patient staging with respect to subject-specific orderings (as done in HEBM) can also be considered when extending DEBM for longitudinal data, where the subject-specific orderings might be estimated with higher confidence.

6.3. Scalability of Event-Based Models

Understanding the progression of several imaging and non-imaging biomarkers after disease onset is important for assessing the severity of the disease. Hence it is desirable to have a model scalable to a large number of biomarkers. FEBM and DEBM are scalable to large number of events, whereas HEBM is not. This was seen in the simulation experiment on varying number of events (Experiment 8), where the errors of FEBM and DEBM increased asymptotically with increasing number of events. The ordering errors of HEBM reached 0.5 for large number of events, which is equivalent to random prediction.

In Experiment 7, we observed that the errors of the EBMs decrease with increasing number of subjects in the dataset. We hence expect FEBM, DEBM and HEBMvv2 to be scalable to a large number of subjects.

The performance of HEBMjh is seen to be consistently worse than FEBMay in Experiment 4. This is in contrast with the findings of Venkatraghavan et al. (2017), where HEBMjh performed better than FEBMay when the number of biomarkers used were 7, while it performed worse when the number of biomarkers used were 42. One of the key differences between the experiment performed in Venkatraghavan et al. (2017) and Experiment 4 is the number of subjects in the simulation dataset. While the previous study considered 509 subjects, Experiment 4 considered 1737 subjects. HEBMjh jointly estimates the subject-specific orderings of all the subjects and the mixture model to represent the biomarkers in different diagnostic groups. We think that while the joint estimation was good for low number of subjects, increasing the number of subjects had an adverse effect on the convergence of the algorithm. Hence HEBMjh is not scalable to a large number of subjects.

We decoupled the mixture model and estimation of subject-specific orderings in $HEBM_{vv2}$ (Experiments 4, 6, 7 and 8). This made HEBM more scalable as it improved the results in Experiment 4 with 1737 subjects, but the decoupling had an adverse effect on the algorithm when the number of subjects was low, as seen in Experiment 7, where $HEBM_{vv2}$ performs worse than $FEBM_{vv2}$ even when the number of subjects was low.

FEBM and HEBM are generative approaches for estimating the central ordering. Our results suggest that HEBM is not very scalable. Although FEBM is scalable, the assumptions made in FEBM are too restrictive for heterogeneous disease such as AD. DEBM is a discriminative approach to event-based modeling, which is both scalable and can robustly estimate central ordering even when the disease is heterogeneous.

6.4. The mixture model

The optimization technique for the Gaussian mixture model that is presented in this paper decouples the optimization of Gaussian parameters and mixing parameters. When the Gaussians of the pre-event and post-event classes are highly overlapping, the optimum mixing parameter changes a lot even for small changes in the Gaussian parameters. By decoupling the optimizations for Gaussian parameters and mixing parameters, we get more stable mixing parameters. This helps in improving the accuracy of all EBMs. This was observed in Experiment 4.

6.5. The importance of good biomarkers

Quality of biomarkers plays a huge role in the accuracy of the EBMs. This was seen in Experiment 8, where the mean error value for 7 biomarkers was considerably higher than the mean error value with the same number of biomarkers in Experiment 5 (for the same Σ_β and Σ_ξ parameters). The observed difference can be explained by the choice of the biomarkers used in those experiments. While the biomarkers chosen in Experiment 8 was at random, the ones chosen in Experiment 5 were the 7 best biomarkers.

6.6. Interpretation of model results on ADNI

Experiment $1(a)$ showed that CSF biomarker ABETA is the first biomarker to become abnormal, followed by MMSE, p-TAU and ADAS13. However, Experiment $3(a)$ showed that in the homogeneous subset of subjects showing signs of typical AD progression (with ABETA positive subjects in MCI and AD, and with ABETA negative CN subjects) p-TAU becomes abnormal before cognitive biomarkers of ADAS13 and MMSE, which is in agreement with Jack's hypothetical model (Jack Jr. et al., 2013). The earlier position of MMSE in Experiment 1 as compared to Experiment 3 can be attributed to the inclusion of SMC subjects as well, who need not necessarily be progressing towards AD. The ordering of p-TAU becoming abnormal before ADAS13 which is then followed by Hippoccampus was also observed by Donohue et al. (2014) in ABETA positive subjects.

ADAS13 and MMSE are seen to become abnormal quite early in the disease progression timeline in Experiments 1 and 3. This is in agreement with other studies on prodromal Alzheimer's Disease (Amieva et al., 2005, 2008). Cognitive biomarkers becoming abnormal before abnormality in Hippocampus and other structural biomarkers, as seen in Experiments 1 and 3, could be due to the fact that the region-based volumes from structural MRI may not be sensitive enough to detect mild structural changes.

The event centers of Hippocampus volume and TAU are quite close to each other in both Experiment $1(a)$ and $3(a)$, which is also in agreement with the current understanding of the disease (de Souza et al., 2012).

Fig. 5 shows that abnormality in the anterior temporal lobe precedes that of the posterior temporal lobe. This was also observed by Schiratti et al. (2015), where the anterior temporal lobe had a higher averaged acceleration factor than the posterior temporal lobe, in a study on AD patients and stable controls.

Nucleus accumbens right and left are the first biomarkers to become abnormal as seen in Fig. 6. This was also observed by Young et al. (2018) in one of the subtypes of AD identified in their work. However, the large standard error of the event centers for the events before ABETA suggests that the exact position of those events are unreliable. Experiment $1(b)$ showed that weak biomarkers (biomarkers excluded in Fig. 5, but included in Fig. 6) could lead to greater uncertainty in event centers. This can be explained by the fact that weak biomarkers are the ones where there is a lot of overlap between the Gaussians of pre-event and post-event classes. Small variation in the sampling population during bootstrapping leads to large changes in the parameters estimated in the mixture modeling step of the algorithm. It also showed that majority of the early structural biomarkers are from Temporal lobe, followed by Central structures, Frontal lobe, Parietal lobe and Occipital lobe.

7. Conclusion

We proposed a new framework for event-based modeling, called discriminative event-based modeling (DEBM), which includes a new optimization strategy for Gaussian mixture modeling, a new paradigm for inferring the mean ordering, a way for estimating the proximity of events in the order to create a disease progression timeline, and a new way of staging patients that uses these relative proximities of events while placing new subjects on the estimated timeline. The source code for DEBM and FEBM was made publicly available online under the GPL 3.0 license: https://github.com/88vikram/pyebm/.

We applied the DEBM framework to a set of 1737 subjects from the baseline ADNI measurement, and also performed an extensive set of simulation experiments verifying the technical validity of DEBM. The experiment on ADNI data illustrated a number of advantages of the new approach. Firstly, we showed that strict quantization of position in ordering of events in the positional variance diagram can sometimes be non-intuitive in terms of inferring actual progression of a disease. Secondly, we showed that the patient staging based on the proposed approach separates CN and AD group of subjects much better than the previous EBM models. Thirdly, we showed that the patient staging can be used to identify individuals at-risk of developing AD as the MCI converters and non-converters were well-separated. Staging patients based on the estimated disease progression timeline can thus make computer-aided diagnosis and prognosis more explainable. The results of these experiments are encouraging and suggest that DEBM is a promising approach to disease progression modeling.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2018.11.024.

References

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J.M., Le Carret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., Dartigues, J.F., 2005. The 9 year cognitive decline before dementia of the alzheimer type: a prospective population-based study. Brain 128, 1093–1101.

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J.M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., Dartigues, J.F., 2008. Prodromal alzheimer's disease: successive emergence of the clinical symptoms. Ann. Neurol. 64, 492–498.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Bron, E.E., Steketee, R.M., Houston, G.C., Oliver, R.A., Achterberg, H.C., Loog, M., van Swieten, J.C., Hammers, A., Niessen, W.J., Smits, M., Klein, S., for the Alzheimer's Disease Neuroimaging Initiative, 2014. Diagnostic classification of arterial spin labeling and structural MRI in presenile early stage dementia. Hum. Brain Mapp. 35, 4916–4931.

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

Fligner, M.A., Verducci, J.S., 1988. Multistage ranking models. J. Am. Stat. Assoc. 83, 892–901.

Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C., Alexander, D.C., 2012. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. Neuroimage 60, 1880–1889.

Gousias, I.S., Rueckert, D., Heckemann, R.A., Dyet, L., Boardman, J.P., Edwards, A.D., Hammers, A., 2008. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. Neuroimage 40, 672–684.

Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum. Brain Mapp. 19, 224–247.

Huang, J., Alexander, D., 2012. Probabilistic event cascades for Alzheimer's disease. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp. 3095–3103

Iturria-Medina, Y., Sotero, R.C., Toussaint, P.J., Mateos-Pérez, J.M., Evans, A.C., of the Alzheimer's Disease Neuroimaging Initiative, 2016. Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. Nat. Commun. 7, 11934.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R., Elliott, C., Masliah, E., Ryan, L., Silverberg, N., 2018. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimer's Dementia 14, 535–562.

Jack Jr., C.R., Knopman, D.S., Jagust, W.J., Petersen, R.C., Weiner, M.W., Aisen, P.S., Shaw, L.M., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Pankratz, V.S., Donohue, M.C., Trojanowski, J.Q., 2013. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 12, 207–216.

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

Kumar, R., Vassilvitskii, S., 2010. Generalized distances between rankings. In: Proceedings of the 19th International Conference on World Wide Web. ACM, New York, NY, USA, pp. 571–580.

Li, D., Iddi, S., Thompson, W.K., Donohue, M.C., for the Alzheimers Disease Neuroimaging Initiative, 2017. Bayesian latent time joint mixed effect models for multicohort longitudinal data. Stat. Methods Med. Res. https://doi.org/10.1177%2F0962280217737566.

Lorenzi, M., Filippone, M., Frisoni, G.B., Alexander, D.C., Ourselin, S., 2017. Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in alzheimer's disease. Neuroimage. https://doi.org/10.1016/j.neuroimage.2017.08.059.

Marinescu, R.V., Oxtoby, N.P., Young, A.L., Bron, E.E., Toga, A.W., Weiner, M.W., Barkhof, F., Fox, N.C., Klein, S., Alexander, D.C., EuroPOND, A.D.N.I., 2018. Tadpole Challenge: Prediction of Longitudinal Evolution in Alzheimer's Disease. Arxiv.

Oxtoby, N.P., Alexander, D.C., 2017. Imaging plus X: multimodal models of neurodegenerative disease. Curr. Opin. Neurol. 30, 371–379.

Prince, M., Wimo, A., Guerchet, M., Ali, G.C., Wu, Y.T., Prina, M., 2015. World Alzheimer's report 2015, the global impact of dementia: an analysis of prevalence, incidence, cost and trends. Alzheimer's Disease Int'l . 4.

Sabuncu, M.R., Bernal-Rusiel, J.L., Reuter, M., Greve, D.N., Fischl, B., 2014. Event time analysis of longitudinal neuroimage data. Neuroimage 97, 9–18.

Schiratti, J.B., Allassonnière, S., Routier, A., Colliot, O., Durrleman, S., 2015. A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In: Ourselin, S., Alexander, D.C., Westin, C.F., Cardoso, M.J. (Eds.), Information Processing in Medical Imaging. Springer International Publishing, Cham, pp. 564–575.

Schmidt-Richberg, A., Ledig, C., Guerrero, R., Molina-Abril, H., Frangi, A., Rueckert, D., of the Alzheimers Disease Neuroimaging Initiative, 2016. Learning biomarker models for progression estimation of Alzheimers disease. PloS One 11, 1–27.

Schott, J.M., Bartlett, J.W., Fox, N.C., Barnes, J., for the Alzheimer's Disease Neuroimaging Initiative Investigators, 2010. Increased brain atrophy rates in cognitively normal older adults with low cerebrospinal fluid Aβ1-42. Ann. Neurol. 68, 825–834.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P.S., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M.Y., Trojanowski, J.Q.a., 2009. Cerebrospinal fluid biomarker signature in alzheimer's disease neuroimaging initiative subjects. Ann. Neurol. 65, 403–413.

de Souza, L.C., Chupin, M., Lamari, F., Jardel, C., Leclercq, D., Colliot, O., Lehéricy, S., Dubois, B., Sarazin, M., 2012. CSF tau markers are correlated with hippocampal volume in Alzheimer's disease. Neurobiol. Aging 33, 1253–1257.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imag. 29, 1310–1320.

Venkatraghavan, V., Bron, E.E., Niessen, W.J., Klein, S., 2017. A discriminative event based model for Alzheimer's disease progression modeling. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.T., Shen, D. (Eds.), Information Processing in Medical Imaging. Springer International Publishing, Cham, pp. 121–133.

Young, A.L., Marinescu, R.V.V., Oxtoby, N.P., Bocchetta, M., Yong, K., Firth, N., Cash, D.M., Thomas, D.L., Dick, K.M., Cardoso, J., van Swieten, J., Borroni, B., Galimberti, D., Masellis, M., Tartaglia, M.C., Rowe, J.B., Graff, C., Tagliavini, F., Frisoni Jr., G.,R.L., Finger, E., Medonậa, A., Sorbi, S., Warren, J.D., Crutch, S., Fox, N.C., Ourselin, S., Schott, J.M., Rohrer, J.D., Alexander, D.C., GENFI, A.D.N.I., 2018. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. Nat. Commun. 9, 4273.

Young, A.L., Oxtoby, N.P., Daga, P., Cash, D.M., Fox, N.C., Ourselin, S., Schott, J.M., Alexander, D.C., 2014. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain 137, 2564–2577.

Young, A.L., Oxtoby, N.P., Huang, J., Marinescu, R.V., Daga, P., Cash, D.M., Fox, N.C., Ourselin, S., Schott, J.M., Alexander, D.C., 2015a. Multiple Orderings of Events in Disease Progression. Springer International Publishing, Cham, pp. 711–722.

Young, A.L., Oxtoby, N.P., Ourselin, S., Schott, J.M., Alexander, D.C., 2015b. A simulation system for biomarker evolution in neurodegenerative disease. Med. Image Anal. 26, 47–56.