





Learning Transition Times in Event Sequences: The Temporal Event-Based Model of Disease Progression

Peter A. Wijeratne^(✉)  and Daniel C. Alexander ,
for the Alzheimer's Disease Neuroimaging Initiative

Centre for Medical Image Computing, Department of Computer Science,
University College London, London, UK
{p.wijeratne,d.alexander}@ucl.ac.uk

Abstract. Progressive diseases worsen over time and can be characterised by sequences of events that correspond to changes in observable features of disease progression. Here we connect ideas from two formerly separate methodologies – event-based and hidden Markov modelling – to derive a new generative model of disease progression: the Temporal Event-Based Model (TEBM). TEBM can uniquely infer the most likely group-level sequence and timing of events (*natural history*) from mixed data types. Moreover, it can infer and predict individual-level trajectories (*prognosis*) even when data are missing, giving it high clinical utility. Here we derive TEBM and provide an inference scheme based on the expectation maximisation algorithm. We use imaging, clinical and biofluid data from the Alzheimer's Disease Neuroimaging Initiative to demonstrate the validity and utility of our model. First, we train TEBM to uncover a new sequence and timing of events in Alzheimer's disease, which are inferred to occur over a period of ~ 17.6 years. Next, we demonstrate the utility of TEBM in predicting clinical progression, and that TEBM provides improved utility over a comparative disease progression model. Finally, we demonstrate that TEBM maintains predictive accuracy with up to 50% missing data. These results support the clinical validity of TEBM and its broader utility in real-world medical applications.

Keywords: Bayesian network · Markov jump process · Disease progression model · Prognosis · Dementia

1 Introduction

Progressive diseases such as Alzheimer's disease (AD) are characterised by monotonic deterioration in functional, cognitive and physical abilities over a period of

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.

years to decades [1]. AD has a long prodromal phase before symptoms become manifest (~ 20 years), which presents an opportunity for therapeutic intervention if individuals can be identified at an early stage in their disease trajectory [2]. Clinical trials for disease-modifying therapies in AD would also benefit from methods that can stratify participants, both in terms of individual-level disease stage and rate of progression [3].

Data-driven models of disease progression can be used to learn hidden information, such as individual-level stage, from observed data [4]. Broadly speaking, disease progression models can be categorised as regression-based or state-space models. Recent examples of regression-based methods include Bayesian mixed-effects models [5,6]; and a time-reparametrised Gaussian Process Progression Model (GPPM) [7]. However, [5,6] require strong assumptions of sigmoidal forms for trajectory dynamics, which may not reflect the true form in the data. The GPPM removes the sigmoidal assumption and allows for non-parametric dynamics, but requires regularisation to avoid over-fitting which introduces additional complexity, and does not learn individual-level rates of progression.

A state-space model of disease progression was proposed by [8], who derived a modified continuous time hidden Markov model (CTHMM), which they used to learn a set of disease states and transition times between these states from electronic health record data. However, CTHMMs fit distribution parameters for each state directly, which increases the number of model parameters that need to be inferred with the number of states. This in turn increases the likelihood of over-fitting [9], which is a well known problem for Markov models when data are sparse and/or the model is complex [10].

An event-based model (EBM) of disease progression was first proposed by [11]. The EBM defines disease progression as a monotonically ordered sequence of binary abnormality events, and as such is essentially a state-space model. Unlike a hidden Markov model, the EBM uses the monotonicity assumption to define a prior form to the distributions generating the data in each hidden state, which simplifies the inference problem. The robustness and predictive utility of the EBM was demonstrated by [12] and extended by [13] to enable both subtype (i.e., multiple sequence) and stage inference (SuStaIn). The simplicity, interpretability and utility of EBM and SuStaIn has made them popular tools for revealing new disease insights [12–20], for validating new features of disease progression (*biomarkers*) [21], and for patient stratification [13,20].

However, EBM and SuStaIn are formulated for cross-sectional data, and hence can only infer the sequence of events but not their transition times, and cannot account for individual-level time series data. Previous work has attempted to address the problem of estimating transition times in an *ad hoc* manner by first fitting an EBM and then correlating its predictions with a separate longitudinal model [16]. However, this approach provides only an approximation of time between events that is confounded by differing model assumptions. As such there is demand for a single unified method that generalises the EBM to accommodate longitudinal data and to simultaneously infer both the order and timing of events.

Here we address this problem, which is long-standing in the field of disease progression modelling [11, 22]. We connect ideas from two formerly separate methodologies – event-based and hidden Markov modelling – to derive a new generative temporal event-based model (TEBM) of disease progression. As a disease progression model, TEBM provides a natural framework to integrate mixed data types, such as imaging and clinical markers, in an informative manner. TEBM therefore has strong clinical utility, as it learns an interpretable group-level model of how mixed biomarkers change over time. Such a model for AD was first hypothesised by [23], and [6, 7, 16] all reported trajectories of various biomarker changes, but TEBM is the first to provide a single unified methodology for learning data-driven sequences and transition times in progressive diseases. As such, this paper has three main contributions.

1. We derive TEBM by generalising the EBM to longitudinal data. TEBM inherits the capabilities of EBM, which can *i*) learn an interpretable sequence of events underlying disease progression; *ii*) learn an individual-level disease stage; *iii*) handle partially missing data (when an individual does not have measurements for every feature). In addition, TEBM can uniquely *iv*) learn transition times; *v*) learn an individual-level probability of progression.
2. We devise a novel algorithm for inference of the TEBM parameters.
3. We apply TEBM to data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) to reveal a new sequence and timing of imaging, clinical and biofluid events in Alzheimer’s disease, and to demonstrate TEBM’s improved utility over a CTHMM, and its performance in the presence of missing data.

2 Theory

2.1 Temporal Event-Based Model

To formulate TEBM, we make three assumptions, namely *i*) monotonic biomarker change; *ii*) a consistent event sequence across the whole sample; and *iii*) Markov (memoryless) stage transitions. We can write the TEBM joint distribution over all variables in a hierarchical Bayesian framework:

$$P(S, \theta, k, Y) = P(S) \cdot P(\theta|S) \cdot P(k|\theta, S) \cdot P(Y|k, \theta, S). \quad (1)$$

Here S is the hidden sequence of events, θ are the distribution parameters generating the data, k is the hidden disease stage, and Y are the observed data. Graphical models of CTHMM, EBM and TEBM are shown in Fig. 1. Note that we have assumed conditional independence of S from k ; that is, the complete set of disease progression stages is independent of the time of observation. Assuming independence between observed features $i = 1, \dots, I$, if a patient $j = 1, \dots, J$ is at latent stage $k_{j,t} = 0, \dots, N$ in the progression model, the likelihood of their data $Y_{j,t}$ observed at time $t = 1, \dots, T_j$ is given by:

$$P(Y_{j,t}|k_{j,t}, \theta, S) = \prod_{i=1}^I P(Y_{i,j,t}|k_{j,t}, \theta_i, S). \quad (2)$$

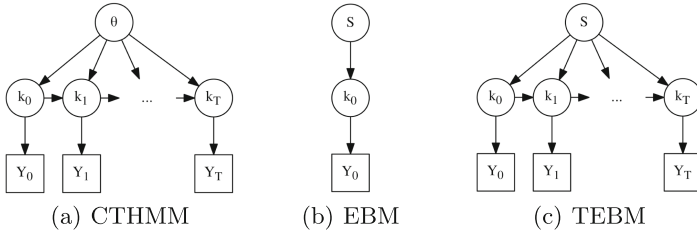


Fig. 1. Graphical models for (a) CTHMM, (b) EBM, and (c) TEBM. Hidden variables are denoted by circles, observations by squares. S : sequence of events; θ : distribution parameters; k : disease stage; Y : observed data; T : observed time.

Here θ_i are the distribution parameters for feature i , defined by a hidden set of events $S = (s(1), \dots, S(N))$. Following [11], we enforce the monotonicity assumption by requiring S to be ordered, which is equivalent to requiring that feature i is monotonic at the group-level. This assumption is necessary to allow snapshots from different individuals to inform on the full event ordering. Next, we assume a Markov jump process [24] between time-points:

$$P(Y_j|k_j, \theta, S) = P(k_{j,t=0}) \prod_{t=1}^{T_j} P(k_{j,t}|k_{j,t-1}) \prod_{t=0}^{T_j} \prod_{i=1}^I P(Y_{i,j,t}|k_{j,t}, \theta_i, S). \quad (3)$$

To obtain an event-based model, we now define prior values for the distribution parameters θ for each stage k in sequence S . Following [12] we choose a two-component Gaussian mixture model to describe the data likelihood:

$$\prod_{i=1}^I P(Y_{i,j,t}|k_{j,t}, \theta_i, S) = \prod_{i=1}^{k_{j,t}} P(Y_{i,j,t}|k_{j,t}, \theta_i^p, S) \prod_{i=k_{j,t}+1}^I P(Y_{i,j,t}|k_{j,t}, \theta_i^c, S). \quad (4)$$

Here $\theta_i^p = [\mu_i^p, \sigma_i^p, w_i^p]$ and $\theta_i^c = [\mu_i^c, \sigma_i^c, w_i^c]$ are the mean, μ , standard deviation, σ , and mixture weights, w , for the patient and control distributions, respectively. Note that these distributions are fit prior to inference, which requires our data to contain labels for patients and controls (see Sect. 3); however, once θ_i^p and θ_i^c have been fit, the model can infer S without any labels. One of the strengths of the mixture model approach is that when feature data are missing, the two probabilities on the RHS of Eq. 4 can simply be set equal.

To obtain the total data likelihood, we marginalize over the hidden stage k and assume independence between measurements from different individuals j (dropping indices j, t in the sum for notational simplicity):

$$P(Y|\theta, S) = \prod_{j=1}^J \left[\sum_{k=0}^N P(k_{j,t=0}) \prod_{t=1}^{T_j} P(k_{j,t}|k_{j,t-1}) \prod_{t=0}^{T_j} \prod_{i=1}^{k_{j,t}} P(Y_{i,j,t}|k_{j,t}, \theta_i^p, S) \prod_{i=k_{j,t}+1}^I P(Y_{i,j,t}|k_{j,t}, \theta_i^c, S) \right]. \quad (5)$$

We can now use Bayes' theorem to obtain the posterior distribution over S . We note that Eq. 5 is the time generalisation of the models presented by [11, 12, 14], and for $T_j = 1$ it reduces to those models.

With this definition made, we now make the usual Markov assumptions [24] to obtain the form of the $N \times N$ dimensional transition generator matrix $Q_{a,b}$:

$$\text{expm}(\Delta Q)_{a,b} = P(k_{j,t} = a | k_{j,t-1} = b, \Delta) \equiv A_{a,b}(\Delta). \quad (6)$$

Here we assume a homogeneous continuous-time process, τ , and a fixed time interval, Δ , that is matrix exponentially distributed, $\Delta \sim \text{expm}(\Delta)$, between stages a, b . Note that as we are only considering Δ constant, $A_{a,b}$ is independent of time, i.e., $A_{a,b}(\Delta) \equiv A_{a,b}$. The N dimensional initial stage probability vector π_a is defined as:

$$\pi_a = P(k_{j,t=0} = a). \quad (7)$$

Finally, the expected duration of each stage (sojourn time), δ_k , is given by:

$$\delta_k = \sum_{\delta=1}^{\infty} \delta P_k(\delta) = \frac{1}{1 - p_{kk}}. \quad (8)$$

Here $P_k(\delta)$ is the probability density function of δ in stage k , and p_{kk} are the diagonal elements of the transition matrix $A_{a,b}$.

2.2 Inference

We aim to learn the sequence \bar{S} , initial probability $\bar{\pi}_a$, and transition matrix $\bar{A}_{a,b}$, that maximise the complete data log likelihood, $\mathcal{L}(\bar{S}, \bar{\pi}, \bar{A}) = \log P(Y|S, \pi, A; \theta)$. As described in Sect. 2, we first obtain θ by fitting Gaussian mixture models to the feature distributions of the patient and control sub-groups. We then use a nested inference scheme based on iteratively optimising the sequence S , and fitting the initial probability π_a and transition matrix $A_{a,b}$, to find a local (possibly global) maximum via a nested application of the Expectation-Maximisation (EM) algorithm. At the first EM step, S is optimised for the current values of the initial probability π'_a and transition matrix $A'_{a,b}$, by permuting the position of every event separately while keeping the others fixed. At the second step, π_a and $A_{a,b}$ are fitted for the current sequence S' using the standard forward-backward algorithm [24]. Here we apply only a single EM pass, as iterative updating of π_a and $A_{a,b}$ can cause over-fitting for sparse [10] and noisy data.

2.3 Staging

After fitting \bar{S} , $\bar{\pi}_a$ and $\bar{A}_{a,b}$, we infer the most likely Markov chain (i.e., trajectory) for each individual using the standard Viterbi algorithm [24]. We can use TE BM to predict individual-level future stage by multiplying the transition matrix, $\bar{A}_{a,b}$, with the posterior probability for the individual at time t , and selecting the maximum likelihood stage:

$$\arg \max_k P(k_{t+1} = b | \bar{S}) = \arg \max_k P(k_t = a | \bar{S}) \cdot \bar{A}_{a,b}. \quad (9)$$

We can also use TE BM to define an individual-level ‘probability of progression’ that leverages the full information from the posterior, which we define as the normalised ratio of the predicted and inferred posteriors:

$$P(k_t, k_{t+1} | \bar{S}) = 1 - \frac{1}{I} \sum_b \frac{P(k_t = a | \bar{S}) \cdot \bar{A}_{a,b}}{P(k_t = a | \bar{S}) \cdot \bar{A}_{a,b} + P(k_t = b | \bar{S})}. \quad (10)$$

For a forward-only transition matrix, Eq. 10 will equal zero if the predicted and inferred posteriors are equal (i.e., zero probability of progression), and non-zero otherwise.

3 Experiments and Results

3.1 Alzheimer’s Disease Data

We use data from the ADNI study, a longitudinal multi-centre observational study of AD [25]. We select 468 participants (119 CN: cognitively normal; 297 MCI: mild cognitive impairment; 29 AD: manifest AD; 23 NA: not available), and three time-points per participant (baseline and follow-ups at 12 and 24 months). Individuals were allowed to have partially missing data at any time-point; we refer to this dataset as ‘Dataset 1’. To facilitate direct comparison between TE BM and CTHMM (the latter of which cannot handle partially missing data by default), we also select a subset of 368 individuals without partially missing data at any time-point, but who can now have different numbers of time-points; we refer to this dataset as ‘Dataset 2’. We train on a mix of 12 clinical, imaging and biofluid features. The clinical data are three cognitive markers: ADAS-13, Rey Auditory Verbal Learning Test (RAVLT) and Mini-Mental State Examination (MMSE). The imaging data are T1-weighted 3T structural magnetic resonance imaging (MRI) scans, post-processed to produce regional volumes using the GIF segmentation tool [26]. We select a set of sub-cortical and cortical regional volumes with reported sensitivity to AD pathology, namely the hippocampus, ventricles, entorhinal, mid-temporal, and fusiform, and the whole brain [27]. The biofluid data are three cerebrospinal fluid markers: amyloid- β_{1-42} (ABETA), phosphorylated tau (PTAU) and total tau (TAU). The ADNI dataset used in this paper is freely available upon registering with an ADNI account and downloading the TADPOLE challenge dataset.

3.2 Model Training

We compare the TEBM and CTHMM models. To ensure fair comparison, we impose a constraint on both models by placing a 2nd order forward-backward prior on the transition matrix. For TEBM, we fit Gaussian mixture models to the distributions of AD (patients) and CN (controls) sub-groups (as in [12]) prior to performing inference, and use 16 start-points for the EM algorithm. For CTHMM, we apply the standard forward-backward algorithm and iterate the likelihood to convergence within 10^{-3} of the total model likelihood. We initialise the CTHMM prior mean and covariance matrices from the training data, using standard k -means and the feature covariance, respectively. Finally, we optimise the number of states in the CTHMM to obtain the maximum likelihood fit. TEBM is implemented and parallelised in Python and is available open source¹. The code takes 3 min to train TEBM with 4 start-points on Dataset 1 using a 4-core 2.7 GHz Intel® Core™ i7-7500U CPU (i.e., 1 start-point per core).

3.3 TEBM Parameters

We train TEBM on Dataset 1 to infer \bar{S} , $\bar{\pi}$ and \bar{A} . Figure 2 shows (a) the Gaussian mixture models for each feature, (b) the initial probability density distribution $\bar{\pi}$; and (c) the event-based transition matrix, \bar{A} , with stages ordered by \bar{S} . The Gaussian mixture models demonstrate smooth transition probabilities between patient and control distributions (denoted by ‘p(event occurred)’), indicating suitable fits. The initial probability distribution is most dense around the earliest stages ($k_{j,t=0} \leq 6$), which reflects the large proportion of CN and MCI individuals in the cohort. The event-based transition matrix is diagonally-dominant, with smooth transitions between stages and predominantly larger forward than backward probabilities, supporting the monotonicity hypothesis.

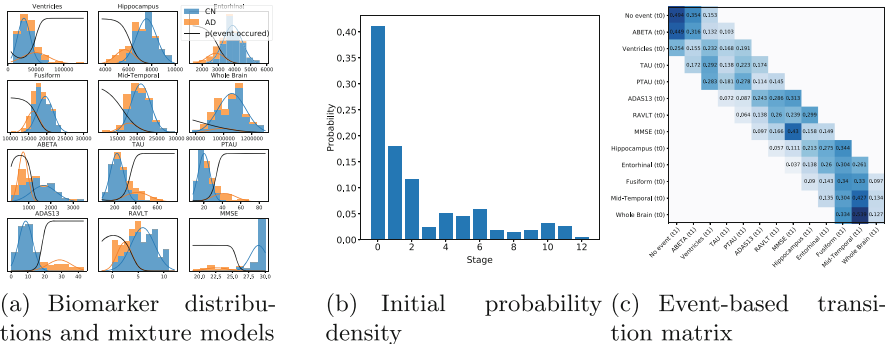


Fig. 2. TEBM parameters inferred from ADNI. (a) Gaussian mixture models fit to distributions of CN (green) and AD (red) groups. (b) Initial probability density, $\bar{\pi}$, inferred by TEBM. (c) Event-based transition matrix, \bar{A} , inferred by TEBM. Events are ordered by the most likely ordering, \bar{S} . (Color figure online)

¹ <https://github.com/pawij/tebm>.

3.4 Alzheimer’s Disease Timeline

We use TEBM to infer the group-level sequence and time between events from Dataset 1, the latter of which is not possible using EBM. Figure 3 shows the corresponding order and timeline of events, and stages for two representative patients estimated by TEBM. This timeline is the first of its type in the field of AD progression modelling, and reveals a chain of observable events occurring over an inferred period of ~ 17.6 years. The ordering largely agrees with previous model-based analyses [12, 16], and TEBM provides additional information on the time between events. Early changes in biofluid measures (ABETA, TAU, PTAU) over a relatively short timescale have been proposed in a recent hypothetical model of AD biomarker trajectories [23], though no actual timing information is reported. We also observe early neurodegeneration (represented here by the ventricles), followed by a chain of cognitive and structural brain volume changes, with change across the whole brain occurring last.

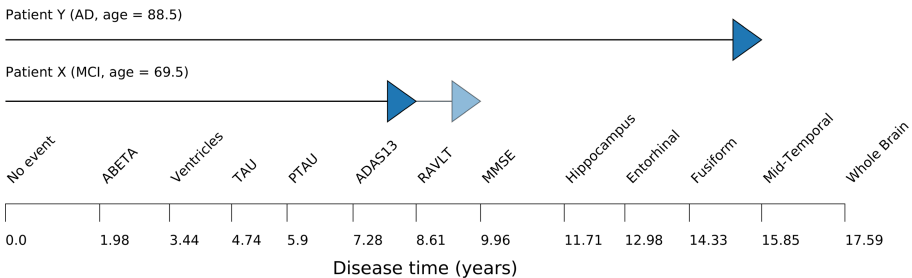


Fig. 3. AD timeline inferred by TEBM. The order of events on the horizontal axis is given by \bar{S} , and the mean time between events is calculated from \bar{A} . Baseline stage (solid arrow) and predicted next stage (shaded arrow, if different to baseline) estimated by TEBM for two example patients from the MCI and AD groups are shown.

3.5 Individual Trajectories

We demonstrate that TEBM can stratify individuals by progression rate and provide a prediction of future stage with uncertainty, which is not possible using EBM. We use TEBM to infer the most likely stage sequence, and predict the most likely next (i.e., unseen) stage over the following year for three individuals. The next most likely unseen stage is predicted according to Eq. 9, with uncertainty estimated by sampling from the posterior. Figure 4 shows three individual trajectories, which were randomly selected from three categories according to their change in stage: stable-stable (no change in stage); progressive-stable (observed increase in stage followed by no predicted change); and progressive-progressive (observed increase in stage followed by predicted increase in stage). These examples highlight the utility of TEBM for clinical applications that aim to stratify by progression rate. If one were to stratify progression rates according

to only observed data, they would be inclined to group individual (a) as stable, (b) as rapidly progressive, and (c) as moderately progressive. However, TEBM predicts that individual (b) remains stable for the following year, while individual (c) increases in stage, making the latter more suitable for observing changes over the following year. As such, TEBM provides additional utility in clinical applications (e.g., prognosis) and clinical trial design (e.g., cohort enrichment).

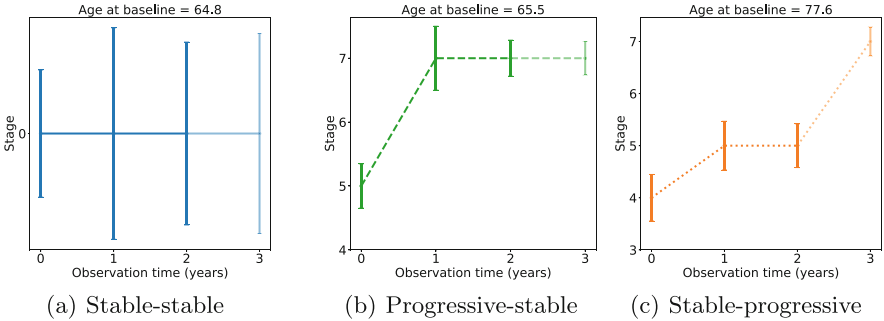


Fig. 4. TEBM individual-level staging. Solid lines represent within-sample inference, and dotted lines represent prediction. Uncertainty was estimated using 100 samples from the posterior at each time-point.

3.6 Prediction of Progression Rate

We now demonstrate TEBM’s ability to predict future rate of progression, which is not possible using EBM. Specifically, we examine the relationship between the rate of change in MMSE – a key cognitive test score used for patient inclusion in AD clinical trials – and the individual-level probability of progression predicted by TEBM (Eq. 10). We train TEBM on 25% of Dataset 1, but use only the baseline measurement from each individual in the test set to predict future progression; this best reflects a baseline clinical trial. Furthermore, we utilise TEBM to select a subset of individuals with a baseline stage = 7, corresponding to abnormal MMSE; this is necessary as individuals at earlier stages are less likely to exhibit abnormal MMSE and hence add noise. We find a nearly significant dependency of MMSE rate of change on TEBM probability of progression (Fig. 5), and in the expected direction, using a linear fixed effects model ($\beta = -20.4$, $p = 0.06$).

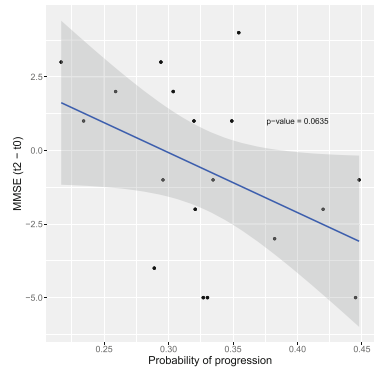


Fig. 5. Dependency of change in MMSE on TEBM probability of progression, for individuals starting at TEBM stage = 7 (abnormal MMSE).

3.7 Comparative Model Performance

We train TEBM and CTHMM to infer individual-level stage sequences and hence compare predictive accuracy on a common task. Specifically, we use baseline stage as a predictor of conversion from CN to MCI, or MCI to AD, over a period of two years. Here predicted converters are defined as people with a stage greater than a threshold stage. We calculate the area under the receiver operating characteristic curve (AU-ROC), and perform 10-fold cross-validation to estimate prediction uncertainty. Table 1(a) shows that TEBM performs substantially better than CTHMM using the same dataset. We also find that TEBM performs better on Dataset 1 than 2, indicating the model benefits when trained on more individuals with more time-points and partially missing data, rather than fewer individuals with fewer time-points and complete data. Note EBM can also infer baseline stage; we find it returns similar performance as TEBM.

Table 1. Model performance for the task of predicting conversion.

(a) TEBM and CTHMM performance for the task of predicting conversion. Only Dataset 2 is reported for CTHMM as it cannot handle partially missing data.

Model	AU-ROC
TEBM (Dataset 1)	0.755 ± 0.12
TEBM (Dataset 2)	0.717 ± 0.15
CTHMM (Dataset 2)	0.489 ± 0.22

(b) TEBM performance for the task of predicting conversion with pre-defined partially missing data, using Dataset 2.

% missing	AU-ROC
25%	0.756 ± 0.12
33%	0.729 ± 0.13
50%	0.723 ± 0.17

3.8 Performance with Missing Data

Finally, we demonstrate TEBM’s ability to handle missing data. We randomly discard 25%, 33% or 50% of the feature data from each individual in Dataset 2 and re-train TEBM. As in Sect. 3.7, we use prediction of conversion as the performance metric, and 10-fold cross-validation. Table 1(b) shows that TEBM maintains consistent performance with up to 50% missing data.

4 Discussion

In this paper we have introduced TEBM, a new generative state-space model of disease progression that combines the strengths of two formerly separate methodologies – *a priori* structure from event-based modelling with temporal information from hidden Markov modelling – to provide a method that can learn transition times in event sequences. The mathematical innovation of our work is to reformulate the EBM in a CTHMM framework (or conversely, the CTHMM in an event-based framework). To our knowledge this is the first such model of its type. We also applied TEBM to reveal a new sequence and timing of key pathological events in AD, and to demonstrate its utility in prediction of conversion and progression at the individual level in the presence of missing data.

TEBM is particularly applicable to clinical trials, where it could be used to inform biomarker and cohort selection criteria. In terms of clinical practice, a key corollary benefit of TEBM’s formulation is that it can infer probabilistic estimates of group- and individual-level progression from datasets with missing data, both in terms of observed features and time-points. This gives it high utility in real-world medical applications, where missing data are present in most (if not all) patient studies, and in clinical applications where resources are scarce and/or it is too expensive to observe a patient multiple times.

Future technical work on TEBM will be focused on relaxing its assumptions, namely *i*) monotonic biomarker change; *ii*) a consistent event sequence across the whole sample; and *iii*) Markov (memoryless) stage transitions. In addition², the assumption of fixed time intervals will be relaxed in future work to accommodate variable intervals between observations, following [28]. Assumption *i*) is both a limitation and a strength: it allows us to simplify our model at the expense of requiring monotonic biomarker change; as shown here, for truly monotonic clinical, imaging and biofluid markers it only provides benefits. However for non-monotonic markers – such as brain connectivity – either the model or data would need to be adapted. Assumptions *ii*) and *iii*) could be relaxed by combining TEBM with (for example) subtype modelling [13] and semi-Markov modelling [29], respectively. In particular, TEBM can be directly integrated into the SuStaIn framework proposed by [13], which would allow us to capture the well-reported heterogeneity in AD and produce timelines such as Fig. 3 for separate subtypes. This opens up the prospect of developing a probabilistic model that can parsimoniously cluster temporal patterns of disease progression, and hence identify clinically-interpretable longitudinal subtypes.

References

1. Masters, C.L., Bateman, R., Blennow, K., et al.: Alzheimer’s disease. *Nat. Rev. Dis. Primers* **1**, 15056 (2015)
2. Dubois, B., Hampel, H., Feldman, H.H., et al.: Preclinical alzheimer’s disease: definition, natural history, and diagnostic criteria. *Alzheimers Dement* **12**(3), 292–323 (2016)
3. Cummings, J., Lee, G., Ritter, A., et al.: Alzheimer’s disease drug development pipeline: 2019. *Alzheimer’s Dement.* **5**, 272–293 (2019)
4. Oxtoby, N.P., Alexander, D.C.: Imaging plus x: multimodal models of neurodegenerative disease. *Curr. Opin. Neurol.* **30**(4), 371–379 (2019)
5. Schiratti, J.B., Allasonnière, S., Colliot, O., et al.: A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *J. Mach. Learn. Res.* **18**, 1–33 (2017)
6. Li, D., Iddi, S., Aisen, P.S., et al.: The relative efficiency of time-to-progression and continuous measures of cognition in presymptomatic Alzheimer’s disease. *Alzheimer’s Dement. Transl. Res. Clin. Interv.* **5**, 308–318 (2019)

² We note that while the requirement of a control sample for fitting the TEBM mixture model distributions could also be deemed a limitation, it is arguably a strength as it allows us to informatively leverage control data; a key issue highlighted by [8].

7. Lorenzi, M., Filippone, M., Frisoni, G.B., et al.: Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in Alzheimer's disease. *NeuroImage* **190**, 56–68 (2019)
8. Wang, X., Sontag, D., Wang, F.: Unsupervised learning of disease progression models. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2014)
9. Lever, J., Krzywinski, M., Altman, N.: Model selection and overfitting. *Nat. Methods* **13**, 703–704 (2016)
10. Ghahramani, Z.: An introduction to hidden Markov models and Bayesian networks. *Int. J. Pattern Recognit. Artif. Intell.* **15**, 9–42 (2001)
11. Fonteijn, H.M., Clarkson, M.J., Modat, M., et al.: An event-based disease progression model and its application to familial alzheimer's disease. *IPMI* **6801**, 748–759 (2011)
12. Young, A.L., Oxtoby, N.P., Daga, P., et al.: A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain* **137**, 2564–2577 (2014)
13. Young, A.L., Marinescu, R.V., Oxtoby, N.P., et al.: Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat. Commun.* **9**, 1–16 (2018)
14. Fonteijn, H.M., Modat, M., Clarkson, M.J., et al.: An event-based model for disease progression and its application in familial alzheimer's disease and huntington's disease. *NeuroImage* **60**, 1880–1889 (2012)
15. Marinescu, R.V., Young, A.L., Oxtoby, N.P., et al.: A data-driven comparison of the progression of brain atrophy in posterior cortical atrophy and alzheimer's disease. *Alzheimer's Dement.* **12**, 401–402 (2016)
16. Oxtoby, N.P., Young, A.L., Cash, D.M., et al.: Data-driven models of dominantly-inherited alzheimer's disease progression. *Brain* **141**, 1529–1544 (2018)
17. Eshaghi, A., Marinescu, R.V., Young, A.L., et al.: Progression of regional grey matter atrophy in multiple sclerosis. *Brain* **141**, 1665–1677 (2018)
18. Firth, N.C., Startin, C.M., Hithersay, R., et al.: Aging related cognitive changes associated with alzheimer's disease in down syndrome. *Ann. Clin. Transl. Neurol.* **5**, 1665–1677 (2018)
19. Wijeratne, P.A., Young, A.L., Oxtoby, N.P., et al.: An image-based model of brain volume biomarker changes in huntington's disease. *Ann. Clin. Transl. Neurol.* **5**, 570–582 (2018)
20. Young, A.L., Bragman, F.J.S., Rangelov, B., et al.: Disease progression modeling in chronic obstructive pulmonary disease. *AJRCCM* **201**(3), 294–302 (2019)
21. Byrne, L.M., Rodrigues, F.B., Johnson, E.B., et al.: Evaluation of mutant huntingtin and neurofilament proteins as potential markers in Huntington's disease. *Sci. Transl. Med.* **10**, eaat7108 (2018)
22. Huang, J., Alexander, D.C.: Probabilistic event cascades for alzheimer's disease. In: *Advances in Neural Information Processing Systems* 25 (2012)
23. Jack, C.R., Holtzman, D.M.: Biomarker modeling of alzheimer's disease. *Neuron* **80**(6), 1347–1358 (2013)
24. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. *IEEE* **77**, 257–286 (1989)
25. Mueller, S.G., Weiner, M.W., Thal, L.J., et al.: The alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. N. Am.* **15**, 869–877 (2005)
26. Cardoso, M.J., Modat, M., Wolz, R., et al.: Geodesic information flows: spatially-variant graphs and their application to segmentation and fusion. *IEEE Trans. Med. Imaging* **34**, 1976–1988 (2015)

27. Frisoni, G.B., Fox, N.C., Jack, C.R., et al.: The clinical use of structural MRI in alzheimer disease. *Nat. Rev. Neurol.* **6**(2), 67–77 (2010)
28. Metzner, P., Horenko, I., Schütte, C.: Generator estimation of markov jump processes based on incomplete observations non-equidistant in time. *Phys. Rev. E. Stat. Nonlin. Soft. Matter Phys.* **76**, 066702 (2007)
29. Alaa, A.M., van der Schaar, M.: A hidden absorbing semi-markov model for informatively censored temporal data: learning and inference. *J. Mach. Learn. Res.* **70**, 60–69 (2018)