A simulation system for biomarker evolution in neurodegenerative disease

Alexandra L. Young [a,*], Neil P. Oxtoby [a], Sebastien Ourselin [a], Jonathan M. Schott [b], Daniel C. Alexander [a], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Centre for Medical Image Computing, Department of Computer Science, University College London, Gower Street, London, WC1E 6BT, United Kingdom
[b] Dementia Research Centre, Institute of Neurology, University College London, 8-11 Queen Square, London, WC1N 3AR, United Kingdom

ARTICLE INFO

ABSTRACT

We present a framework for simulating cross-sectional or longitudinal biomarker data sets from neurodegenerative disease cohorts that reflect the temporal evolution of the disease and population diversity. The simulation system provides a mechanism for evaluating the performance of data-driven models of disease progression, which bring together biomarker measurements from large cross-sectional (or short term longitudinal) cohorts to recover the average population-wide dynamics. We demonstrate the use of the simulation framework in two different ways. First, to evaluate the performance of the Event Based Model (EBM) for recovering biomarker abnormality orderings from cross-sectional datasets. Second, to evaluate the performance of a differential equation model (DEM) for recovering biomarker abnormality trajectories from short-term longitudinal datasets. Results highlight several important considerations when applying data-driven models to sporadic disease datasets as well as key areas for future work. The system reveals several important insights into the behaviour of each model. For example, the EBM is robust to noise on the underlying biomarker trajectory parameters, under-sampling of the underlying disease time course and outliers who follow alternative event sequences. However, the EBM is sensitive to accurate estimation of the distribution of normal and abnormal biomarker measurements. In contrast, we find that the DEM is sensitive to noise on the biomarker trajectory parameters, resulting in an over estimation of the time taken for biomarker trajectories to go from normal to abnormal. This over estimate is approximately twice as long as the actual transition time of the trajectory for the expected noise level in neurodegenerative disease datasets. This simulation framework is equally applicable to a range of other models and longitudinal analysis techniques.

1. Introduction

Neurodegenerative diseases, such as Alzheimer's disease (AD), Huntington's disease (HD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS), present increasing social and economic costs. Determining the sequence and evolution of the symptoms and pathologies of neurodegenerative diseases will enable pre-symptomatic and differential diagnosis, and treatment monitoring for drug trials. Biomarkers have been developed that allow the presence and progression of these pathologies to be measured in vivo. Such biomarkers include cerebrospinal fluid (CSF) measures of proteins implicated in disease pathogenesis, structural magnetic resonance imaging (MRI) measures of regional volume changes, positron emission tomography (PET) measures of hypometabolism or abnormal protein deposition, and cognitive test scores. Recent multi-centre collaborations, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) for AD, the Parkinson Progression Marker Initiative (PPMI) for PD, and the Track-HD study for HD, collect a diverse set of biomarker data from large cohorts. However, analysing longitudinal biomarker trends in such datasets is difficult due to inherent disease heterogeneity and the long disease time course (thought to be over a decade in some cases), which means that even so-called longitudinal data is almost cross-sectional with respect to the full disease duration. As a result, understanding of the longitudinal evolution of biomarkers in neurodegenerative diseases remains largely hypothetical (Aisen et al., 2010; Frisoni et al., 2010; Jack et al., 2010).

* Corresponding author. Tel.: +44 20 7679 0221.
E-mail address: alexandra.young.11@ucl.ac.uk (A.L. Young).

In AD for example, the most well validated biomarkers are amyloid PET imaging (Clark et al., 2011; Klunk et al., 2004) and CSF Aβ_{1-42} (Blennow and Hampel, 2003) to measure brain amyloid pathology; CSF total tau (t-tau) and phosphorylated tau (p-tau) (Blennow and Hampel, 2003) to measure neurofibrillary tangle (NFT) deposition and neuroaxonal damage; FDG-PET imaging (Herholz, 2012) to measure brain hypometabolism; volumetric measures from MRI (Fox and Schott, 2004) to measure neurodegeneration; and a range of cognitive tests to measure memory loss and other cognitive deficits. Hypothetical models of AD describe a distinct sequence in which these biomarkers become abnormal (Jack et al., 2010). Jack et al. (Jack et al., 2010) propose a model in which biomarkers evolve sigmoidally over time with amyloid plaque biomarkers such as CSF Aβ_{1-42} and amyloid PET preceding NFT deposition and neuroaxonal damage markers such as CSF p-tau, CSF t-tau. These in turn become abnormal prior to FDG-PET hypometabolism and grey-matter atrophy measured on MRI, before finally memory and cognition are affected, as measured using cognitive test scores. More recently, Jack et al. (Jack and Holtzman, 2013; Jack et al., 2013a) have revised this model to account for the possibilities that (1) NFT build up and neurodegeneration precede amyloid plaque development in some cases, and (2) subjects have mixed pathology. However, these models await empirical validation from measured datasets.

Data-driven models of disease progression allow longitudinal trends to be reconstructed from cross-sectional or short-term longitudinal datasets. Basic techniques to analyse biomarker trajectories involve staging subjects and then comparing biomarker levels across different disease stages (Bateman et al., 2012; Caroli and Frisoni, 2010; Förster et al., 2012; Jack et al., 2011, 2012; Landau et al., 2012; Lo et al., 2011; Sabuncu et al., 2011; Schuff et al., 2012). This limits the temporal resolution of the model to the accuracy of the patient staging. Patient staging techniques include clinical diagnoses (Förster et al., 2012; Jack et al., 2011; Landau et al., 2012; Lo et al., 2011), which typically comprise just three stages: cognitively normal (CN), mild cognitive impairment (MCI) and clinical AD; cognitive test scores (Caroli and Frisoni, 2010; Jack et al., 2012; Sabuncu et al., 2011), which suffer from floor and ceiling effects; age (Schuff et al., 2012), for which disease stage varies widely amongst subjects; and predicted age of onset based on parents age of onset (Bateman et al., 2012), which can only be established for genetic disease subtypes and is a crude estimate of actual onset. Data-driven models do not require prior knowledge of the stage of a patient along the disease, allowing the reconstruction of a much more fine-grained picture of disease progression. Differential equation models (DEM) (Jack et al., 2013b; Oxtoby et al., 2014; Sabuncu et al., 2011; Villemagne et al., 2013) model short-term longitudinal patient data as a differential cross-section of a common longitudinal biomarker trajectory; the overall biomarker trajectory is obtained by integrating all of the subject's differential cross-sections. Jedynak et al. (Jedynak et al., 2012) and Donohue et al. (Donohue et al., 2014) make a similar set of assumptions to formulate their disease progression models, again modelling each subject's biomarker data as a snapshot of a common progression curve, but further allowing for variation in individual progression rates. Another data-driven model, the Event Based Model (EBM) (Fonteijn et al., 2012; Young et al., 2014) considers disease progression as a sequence of events at which biomarkers become abnormal, thereby allowing direct determination of biomarker ordering from entirely cross-sectional data.

However, such data-driven models typically depend on idealised assumptions about the data that they are modelling. First, that all subjects follow the same progression pattern. This is not true in general as large cross-sectional datasets will contain subjects who have different disease subtypes, mixed pathology, have been misdiagnosed, are yet to develop other diseases, or who are aging healthily. Such outliers are particularly prevalent in pre-symptomatic populations where the diagnostic outcome is unknown. Second, a set of parameters that define normal and abnormal biomarker levels. This is

difficult to determine due to the high proportions of pre-symptomatic subjects in typical control populations (for example, a significant proportion of cognitively normal elderly subjects have been found to have biomarker changes consistent with AD (Rowe et al., 2010; Schott et al., 2010)), and misdiagnosis in diseased populations. Third, that the underlying disease time course is well sampled. In reality, pre-symptomatic subjects may not go on to develop the neurodegenerative disease being investigated and therefore the early disease stages might be under sampled or misrepresented, and diseased subjects may not be representative of the very late disease stages where the population thins and severe illness can make data hard to collect.

Here we present a simulation system to generate synthetic biomarker datasets that represent the heterogeneity of sporadic neurodegenerative diseases. Although still based on a model of disease progression, it encapsulates many more variables than the simpler models that are parsimonious enough to fit to current data sets. Thus it provides a platform to evaluate the effect of more brutal simplifications necessary to obtain robust fitting results from working models. Here, we demonstrate this simulation system by evaluating the performance of the EBM and a DEM in determining the sequence of biomarker abnormality from simulated data.

2. Methods

2.1. Generative model of data

We assume the following generative model of sporadic disease datasets. A set of subjects with: follow-up time points, f, from a follow-up distribution $P(f)$; disease subtypes, s, from a subtype distribution $P(s)$; each set of follow-up time points correspond to a set of time points along the disease, t, from a time point distribution $P(t|f)$. At each time point a subset, e, of the biomarkers included in the study are measured for a particular subject according to a biomarker collection distribution $P(e|f)$, i.e. a subset of the biomarkers included in the study are measured in the subset of subjects that have a particular follow-up visit. Each subject has a set of biomarker measurements, x_e, at each time point $t \in t$. The collected biomarker measurements are simulated from a trajectory evolution function $z(t, \theta)$ with parameters θ from a trajectory parameter distribution $P(\theta|s, t)$, and measurement noise ε perturbation from a measurement noise distribution $P(\varepsilon)$. The data for each collected biomarker for each patient for each time point is then $x = z(t, \theta) + \varepsilon$; biomarkers that are not collected are recorded as missing data. Each subject is given a particular diagnosis d from a diagnosis distribution $P(d|x_e)$.

2.2. Simulating sporadic AD

2.2.1. ADNI dataset

We used data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to guide some of the settings for the simulations. The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organisations, as a $60 million, 5-year public–private partnership. For up-to-date information see http://www.adni-info.org.

We downloaded baseline and follow-up data from all subjects in ADNI-1 giving a set of 819 subjects (229 cognitively normal, 398 mild cognitive impairment, 192 Alzheimer's disease). We included the following set of biomarker data: CSF Aβ_{1-42}, CSF t-tau, CSF p-tau; the Mini Mental State Examination (MMSE) (McKhann et al., 1984); baseline MRI volumes of whole brain, hippocampus and ventricles; FDG-PET. MRI volumes were corrected for differences in head size by regressing against total intracranial volume (TIV). FDG-PET uptake values were averaged over the angular gyrus, inferior temporal gyrus, and posterior cingulate gyrus. For simplicity we only model baseline

CSF, as modelling longitudinal CSF requires a new set of measurements for each time point to be modelled, as all the CSF measurements are re-processed once a new follow up is completed.

2.2.2. Generic AD simulation model

We start from a generic model of AD, adapting the settings to perform a stability analysis of the EBM and DEM.

Our generic model of AD is based on the following assumptions:

- At baseline a set of time points t_0 are sampled from a uniform baseline time point distribution $P(t_0) = Unif(0, t_r)$, where t_r is the range of the initial subject time points.
- The follow-up time points are sampled sequentially from a set of possible follow-up times under the assumption that a proportion, r_f, of subjects drop out per year.
- The time points are assumed to follow a normal distribution centred around the time after follow-up, i.e. $P(t|f) = Norm(f + t_0, \sigma_t)$.
- The subtype distribution is $P(s) = Cat(p_{s,m})$, where p_s is the probability that a subject is assigned subtype s. By default there is only one disease subtype and so all subjects follow the typical AD set of biomarker trajectories. Alternative subtypes are used to simulate subjects that do not follow the typical AD sequence of biomarker abnormality, e.g. subjects with other neurodegenerative diseases or who are aging normally.
- The collected subset of biomarkers, e, is sampled sequentially for the available time points by modelling an initial proportion of subjects, p_e, in which the biomarker is collected and a drop out rate per year, r_e, i.e. of the subjects that remain in the study, only a proportion of these have a measurement for a particular biomarker.
- The trajectory evolution function is sigmoidal, as has been hypothesised by Jack et al. (Jack et al., 2010), with parameters $\theta = (a, r, c, g)$, where a is the trajectory minimum, r the range (difference between trajectory maximum and minimum value), c the centre point and g the gradient. To make the magnitude of the gradient a more intuitive quantity, we re-parameterise g so that it is the biomarker 'transition time'. We define this as $\tau = 4/g$, i.e. it is the time taken for the tangent to the sigmoid at the centre point, c, to transition from the minimum biomarker value, a, to the maximum biomarker value, $a + r$. Hence, we have

$$z(t, \theta) = a + \frac{r}{1 + \exp\left(-\frac{4}{\tau}(t - c)\right)},$$

with parameters $\theta = (a, r, c, \tau)$.
- The trajectory parameters are normally distributed according to the trajectory parameter distributions:
  - $P(a) = Norm(\mu_a, \Sigma_a)$
  - $P(r) = Norm(\mu_r, \Sigma_r)$
  - $P(c|s) = Norm(\mu_{c,s}, \Sigma_c)$
  - $P(\tau) = Norm(\mu_\tau, \Sigma_\tau)$

The parameters μ_a, μ_r, $\mu_{c,s}$, μ_τ are the trajectory parameter means; Σ_a, Σ_r, Σ_c, Σ_τ are the inter-subject covariances of the trajectory parameters.

- The measurement noise distribution follows a normal distribution with mean 0 and standard deviation υ, i.e. $P(\varepsilon) = Norm(0, \upsilon)$
- There are three diagnoses, $d = \{CN, MCI, AD\}$, as there are in ADNI, that follow a categorical distribution: $P(d|\mathbf{x}) = Cat(p_{CN}, p_{MCI}, p_{AD})$, where p_{CN} is the probability that a subject is assigned a CN diagnosis, p_{MCI} is the probability that a subject is assigned a MCI diagnosis, and p_{AD} is the probability that a subject is assigned an AD diagnosis. The probability of each diagnosis p_d is evaluated using each subjects biomarker data, $\mathbf{x}$, as $p_d \propto \prod_{i=1}^I Norm(x_i, \mu_{d,i}, \sigma_{d,i})$.

We tested the agreement between data sets generated using the default parameter values and ADNI by calculating the Bhattacharyya coefficient (Bhattacharyya, 1943), *BC*, between simulated data sets and data from ADNI. The Bhattacharyya coefficient measures the similarity between two probability distributions, ranging from 0 to 1, where a Bhattacharyya coefficient of 0 corresponds to no overlap.

$$D_B = \frac{1}{8}(\mu_{ADNI} - \mu_{simulated})^T \Sigma^{-1}(\mu_{ADNI} - \mu_{simulated})$$

$$+ \frac{1}{2}\ln\left(\frac{\det(\Sigma)}{\sqrt{\det(\Sigma_{ADNI})\det(\Sigma_{simulated})}}\right)$$

where

$$\Sigma = \frac{\Sigma_{ADNI} + \Sigma_{simulated}}{2}$$

The Bhattacharyya coefficient, *BC*, is $BC = \exp(-D_B)$.

We find that data sets generated using the default parameters values show good agreement with the ADNI data set, giving an average Bhattacharyya coefficient across 25 sample data sets of 0.99 when considering the biomarkers to be independent, and 0.83 when considering the dependence between biomarkers.

2.2.3. Datasets for EBM stability analysis

We use our generic AD model to perform a set of simulations to assess how robust the EBM and DEM are to different choices of parameters.

For these experiments we assume the following set of default parameters (Table 1, Fig. 1). These default parameters are intended as an idealised basis for the stability analysis, from which each parameter can be varied individually so as to explore the robustness of the models to variations in a particular parameter, independently of other effects. For each experiment we generate synthetic datasets that have 800 subjects and the biomarker set: CSF A$\beta_{1\text{-}42}$, CSF t-tau, CSF p-tau, FDG, MMSE, hippocampal volume, brain volume, ventricular volume.

- Baseline time points: $t_r = 20$, i.e. there is a range of 20 years in which a subject's baseline visit might lie.
- Follow-ups: There are 11 possible follow-up times, as there are in ADNI, at 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, 8, and 9 years from baseline. The drop out rate per year, r_f, is estimated from ADNI as 10%.
- Time points: the standard deviation, σ_t, of the actual time at which each follow-up is taken is approximated from ADNI as $\sigma_t = 0.05$ for $t = 0.5$, 1, 1.5, and $\sigma_t = 0.1$ for $t \geq 2$ (at baseline $\sigma_t = 0$).
- Subtype: $p_{s=1} = 1$, i.e. all subjects have the same disease subtype by default.
- Biomarker collection: The proportion of subjects in which each biomarker is collected at baseline, p_e, and the drop out rate per year, r_e, are estimated from ADNI as $p_e = 100\%$, 85%, 55%, 50%, and $r_e = 0\%$, 10%, 5%, 100% for cognitive test scores, MRI volumes, FDG-PET hypometabolism, and CSF levels respectively (i.e. only baseline CSF modelled).
- Trajectory parameters: μ_a is estimated from the mean biomarker value in cognitively normal subjects from ADNI (Table 1). μ_r is estimated as the difference between the mean biomarker value in Alzheimer's disease and cognitively normal subjects from ADNI (Table 1). For CSF A$\beta_{1\text{-}42}$ we use only amyloid negative cognitively normal subjects, and amyloid positive Alzheimer's disease subjects (amyloid positive is defined as CSF A$\beta_{1\text{-}42} < 192$ pg/ml). We chose settings for μ_c so that the biomarkers become abnormal in the order: Abeta, p-tau, t-tau, FDG-PET, hippocampal volume, MMSE, ventricles, whole brain volume (Table 1). We set μ_τ to 5 years for all biomarkers (Table 1).
- Trajectory inter-subject covariance: Σ_a is estimated from the set of 28 amyloid negative cognitively normal ADNI subjects at baseline that have measurements for all biomarkers (Table 1), we remove the contribution of measurement noise by subtracting the

Table 1

Default parameter values for the EBM. μ_a = mean trajectory minimum, μ_r = mean trajectory range, μ_c = mean trajectory centre point (years), μ_τ = mean trajectory transition time (years). $\upsilon^2\mu_r^2$ = biomarker measurement variance, normalised by the range of the trajectory. $\Sigma_a/\mu_r\mu_r^T$ = covariance of the trajectory minimum, normalised by the range of the trajectory. All other covariance matrices are set to 0 by default.

		Abeta	p-tau	t-tau	FDG	Hippo	MMSE	Vents	Brain
μ_a		243	20	63	6.39	903	29.1	−7400	34300
μ_r		−110	23	63	−1.01	−1630	−5.78	13600	−64000
μ_c		5	6	7	8	10	12	14	15
μ_τ		5	5	5	5	5	5	5	5
υ^2/μ_r^2		0	0	0	0.030	0.006	0.013	0.004	0.022
$\Sigma_a/\mu_r\mu_r^T$	Abeta	0.058	0.001	−0.004	0.014	−0.001	0.002	0.013	0.007
	p-tau		0.150	0.126	0.041	−0.069	−0.015	0.048	−0.106
	t-tau			0.178	0.031	−0.084	−0.003	−0.013	−0.068
	FDG				0.308	0.064	0.008	−0.191	0.109
	Hippo					0.198	−0.007	−0.320	0.326
	MMSE						0.055	0.009	−0.031
	Vents							1.519	−0.896
	Brain								0.960

estimated measurement variance level (see next bullet point). By default we set $\Sigma_r = 0$, $\Sigma_c = 0$, $\Sigma_\tau = 0$.

- Measurement noise: We estimate the measurement noise level υ for each biomarker using baseline and 6 month follow-up measurements in cognitively normal subjects from ADNI (Table 1) under the assumption that fluctuations in controls over a 6 month period are representative of measurement noise. For CSF we only have baseline measurements and so we set the measurement noise to 0, i.e. we model the variance as being purely inter-subject covariance rather than removing the contribution of measurement noise as we do for the other biomarkers.
- Diagnosis: We estimate μ_d and σ_d for each diagnostic group using the available data for each biomarker from ADNI.

In the experiments we vary each of the following parameters of the simulations in turn, and set the rest of the parameters to their default value, generating 25 synthetic datasets for each new parameter value, and fitting the EBM to each sample dataset.

2.2.3. Datasets for DEM stability analysis

For the DEM experiments we initially fitted a DEM to each biomarker using the default simulation settings for the EBM. However, these experiments show that the DEM does not perform well for the levels of noise estimated from ADNI. We therefore simplify the default settings for the DEM to allow us to characterise the types of noise the DEM is most sensitive to. By default we instead generate a single idealised (zero noise) biomarker trajectory with the following settings.

- Biomarker collection: $p_e = 100\%$, $r_e = 0\%$.
- Trajectory parameters: $\mu_a = 0$, $\mu_r = 1$, $\mu_c = 10$, $\mu_\tau = 5$.
- Trajectory inter-subject covariance: $\Sigma_a = 0$, $\Sigma_r = 0$, $\Sigma_c = 0$, $\Sigma_\tau = 0$.
- Measurement noise: $\upsilon = 0$.

In the experiments we again vary each of the following parameters of the simulations in turn, and set the rest of the parameters to their default value, generating 25 synthetic datasets for each new parameter value, and fitting the DEM to each sample dataset.

2.3. The Event Based Model

The Event Based Model (EBM) (Fonteijn et al., 2012) considers disease progression as a sequence of events at which biomarkers transition from a normal level, i.e. as seen in healthy controls, to an abnormal level, i.e. as seen in AD subjects. The maximum likelihood (ML)

ordering of these events can be determined by finding the sequence S that maximises the data likelihood

$$P(X|S) = \prod_{j=1}^{J}\left[\sum_{k=0}^{I}\left(P(k)\prod_{i=1}^{k}P(x_{ij}|E_i)\prod_{i=k+1}^{I}P(x_{ij}|\neg E_i)\right)\right]$$

Here, E_i, $i = 1 \ldots I$, are events, whose occurrence is informed by the corresponding measurements x_{ij} of biomarker i in subject j, $j = 1 \ldots J$ via the biomarker distributions: the likelihood that an event has occurred and thus the corresponding biomarker measurement x_{ij} is abnormal, $P(x_{ij}|E_i)$, or has yet to occur and so the corresponding biomarker measurement is normal, $P(x_{ij}|\neg E_i)$. $P(k)$ is the prior likelihood of being at stage k, where events $E_1, \ldots, E_k$ have occurred, and events $E_{k+1}, \ldots, E_K$ have yet to occur. We assume no prior knowledge of disease stage by choosing the prior $P(k)$ to be uniform. We fit a mixture of normal distributions to determine the mean, μ_E and $\mu_{\neg E}$, and standard deviation, σ_E and $\sigma_{\neg E}$, of the biomarker distributions $P(x|E)$, and $P(x|\neg E)$. To guide the fitting in cases where the biomarker distributions overlap significantly, we constrain the parameters so that the standard deviation of each distribution is less than or equal to the standard deviation of biomarker measurements in the AD and control (CN) population respectively. For missing biomarker values we impute the value of x such that $P(x|E) = P(x|\neg E)$.

2.4. Differential equation model

We fit the DEM to each biomarker separately using a similar technique to Villemagne et al. (Villemagne et al., 2013). We first calculate the rate of change in each subject by fitting a least-square linear regression to the first three available time points for each participant. For each simulation we compared fitting a linear model to the first three available time points (baseline, 0.5 years and 1 year) with fitting a linear model to the three or more available time points (up to a maximum of 9 years). We found that fitting the linear model to the first three available time points produced trajectories with the least error for all experiments, and so we only present results for fitting to the first three available time points. We fit a quadratic differential equation model (representative of the sigmoidal biomarker dynamics modelled in the simulations) to the mean biomarker value of each subject (x_i), and rate of change of each subject ($\frac{\delta x_i}{\delta t}$) estimated from the linear model, i.e. we optimise for A, B and C over all subjects i such that:

$$\frac{\partial x_i}{\partial t} = Ax_i{}^2 + Bx_i + C$$
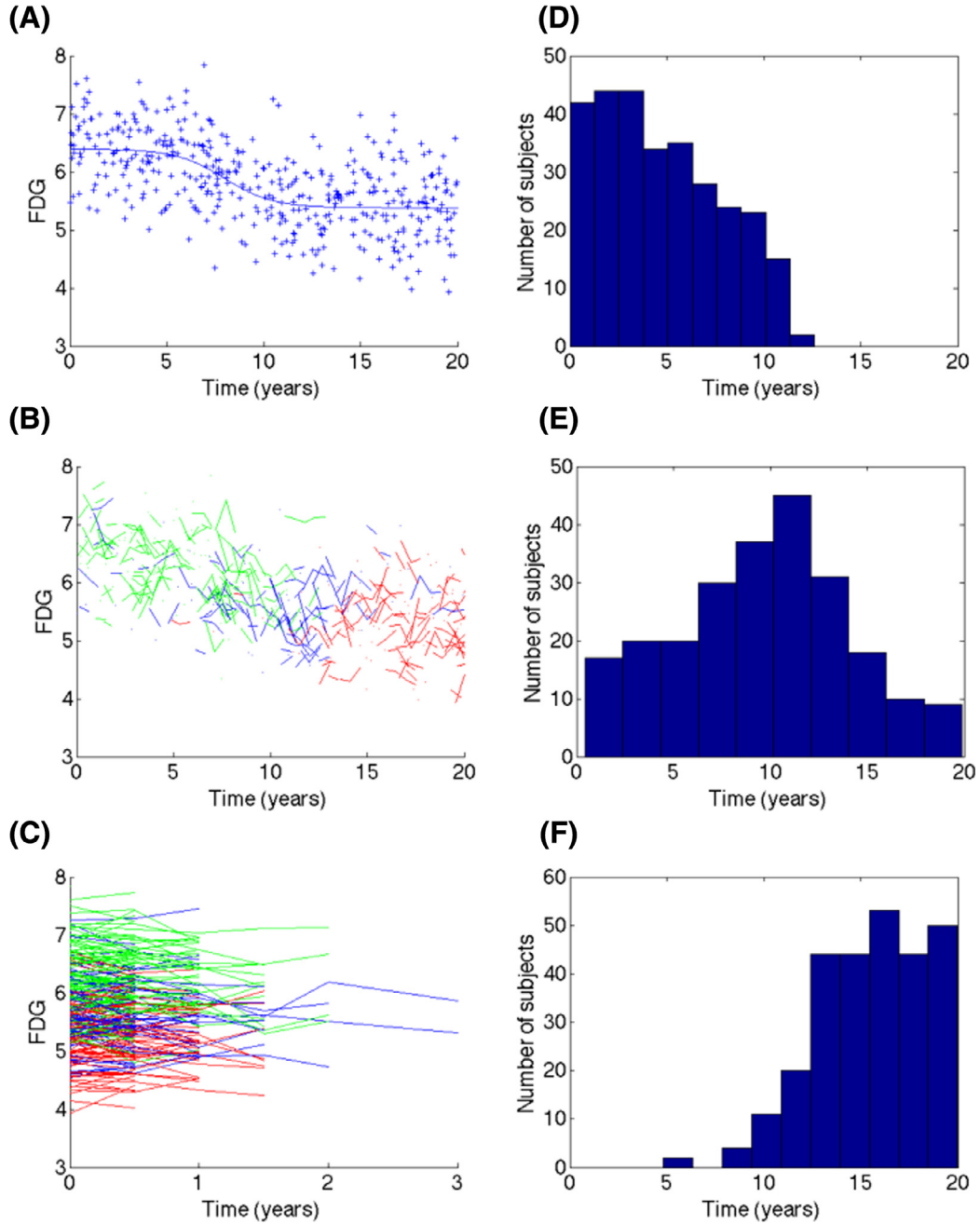
Fig. 1. Example simulated FDG-PET, subject diagnosis, and time point data generated using default parameter values for the EBM (Table 1). (A) Simulated baseline FDG-PET data (curve shows mean trajectory); (B) simulated FDG-PET follow-up data for cognitively normal subjects (green), mild cognitive impairment subjects (blue), and Alzheimer's disease subjects (red); (C) same as (B) but plotted against follow up time rather than time point along the disease. (D)–(F) Histogram of the number of (D) cognitively normal, (E) mild cognitive impairment and (F) Alzheimer's disease subjects, at each time point at baseline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

We then integrate this quadratic differential equation model to get the average trajectory across the population:

$$x(t) = \frac{\sqrt{4AC - B^2} \tan\left(\frac{1}{2}(k+t)\sqrt{4AC - B^2}\right) - B}{2A}$$

where k is an unknown constant to be specified by choosing an initial condition.

2.5. Evaluation metrics

2.5.1. EBM biomarker distribution parameters
Defining a ground truth for the biomarker distribution parameters when the biomarker trajectories are not binary is not straight-forward, requiring the portion of the biomarker trajectory belonging to the 'normal' and 'abnormal' biomarker distribution to be defined. However, to explore the effect of the accuracy of the biomarker distribution on the estimation of the event sequence we ran each of the experiments for two settings: one where we estimated the biomarker distributions and another where these were fixed. We fix the biomarker distributions so that $\mu_{\neg E} = \mu_a$, $\mu_E = \mu_a + \mu_r$, $\sigma_{\neg E} = \sigma_E = \sqrt{\text{diag}(\Sigma_a) + \upsilon^2}$, where $\text{diag}(\Sigma)$ is the diagonal of the covariance matrix Σ. $\mu_{a,m}$ and $\mu_{b,m}$ are calculated from the average subject demographics. Whilst these may not be the 'true' biomarker distribution parameters, they give us an idea of how the EBM behaves for a reasonable setting of the biomarker distribution parameters.
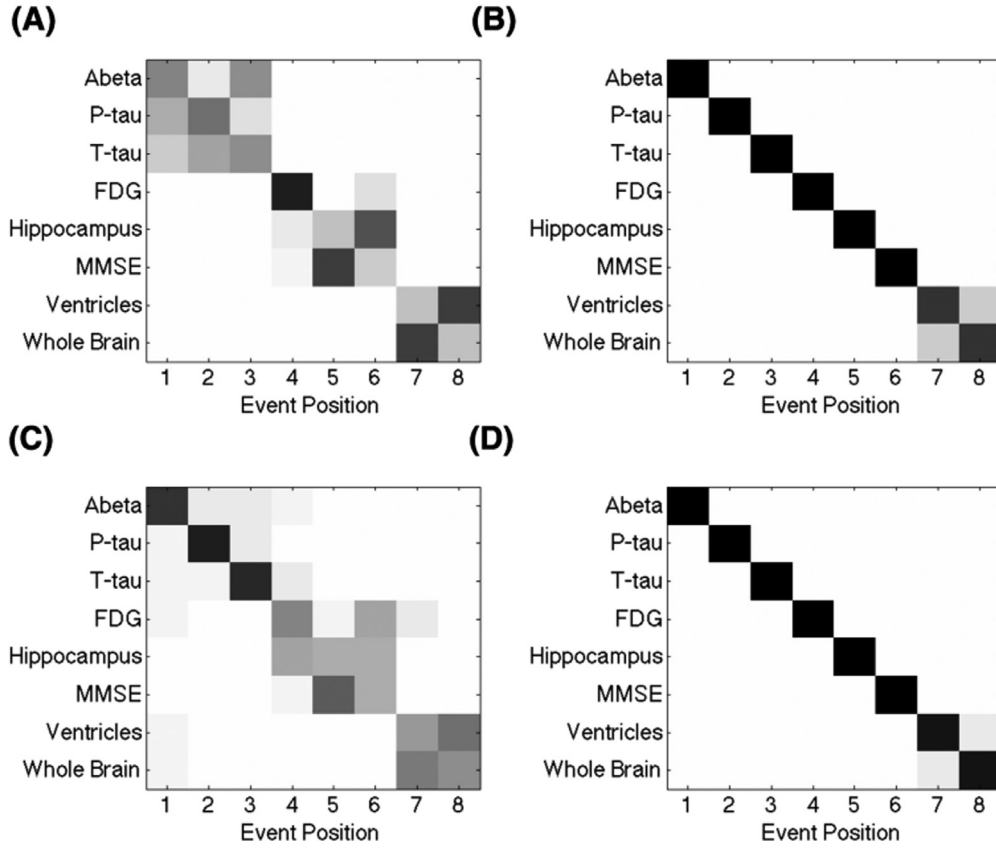
Fig. 2. Results of applying the EBM to synthetic data with missing values (A)–(B) and without missing values (C)–(D) generated using the default parameters. In (A) and (C) the event distributions are estimated, and in (B) and (D) the event distributions are fixed.

2.5.2. EBM Kendall's tau distance

A key outcome measure we are interested in is the model's ability to recover the ML sequence from the simulated data. This can be evaluated by measuring the Kendall's tau distance (a measure of the similarity of two sequences) between the recovered sequence and a ground truth event sequence. The Kendall's tau distance is the total number of pairwise disagreements between two sequences, π and π_0:

$$d_K(\pi, \pi_0) = \sum_{l \prec_\pi j} 1_{[j \prec_{\pi_0} l]}$$

where π and π_0 are permutations and $l \prec_\pi j$ means that l precedes j in the permutation π. Here we use the normalised Kendall's tau distance, i.e. we divide by the maximum distance, which is the total number of possible pairs: $\frac{n!}{2!(n-2)!}$, where n is the number of events in the sequence.

2.5.3. EBM positional variance diagrams

We use positional variance diagrams (PVDs) to look at the location of variations in the ML sequence (rather than just the extent, which we measure using the Kendall's tau distance) for each simulation. Each entry of the positional variance diagram is the proportion of samples in which a particular event appears at that position in the ML sequence.

2.5.4. DEM transition time

We compare the transition time of the simulated trajectories across the population with the ground truth simulation setting of the transition time, μ_τ.

3. Results

3.1. EBM stability analysis

We performed a stability analysis of the EBM to test how robust the model is to different types of heterogeneity that are likely to exist in sporadic AD datasets.

3.1.1. Default parameter values

Fitting the EBM to datasets generated using the default parameter values (Fig. 2) gives a Kendall's tau distance of 0.11 ± 0.05 when estimating the event distributions from the data, and 0.01 ± 0.01 for fixed event distributions. Repeating this experiment without any missing biomarker values only slightly improves the Kendall's tau distance despite the increase in the number of data points; 0.10 ± 0.09 for estimated event distributions, and 0.00 ± 0.01 for fixed event distributions, showing that imputing the data such that $P(x|E) = P(x|\neg E)$ (see Section 2.3) is a valid technique for fitting the EBM to data with missing biomarker values. It is worth noting that in all the simulations the inter-subject variation setting, which is estimated from ADNI, may be larger than the actual level of inter-subject variation. This is because cognitively normal subjects in ADNI may originate from a range of underlying time points along the biomarker trajectories.

3.1.2. Experiment 1: Noise levels

We ran five simulations to look at the effect of different levels of: (A) measurement noise υ, (B) inter-subject covariance of the trajectory minimum Σ_a, (C) inter-subject covariance of the trajectory range Σ_r, (D) inter-subject covariance of the trajectory centre $\Sigma_{c,s}$, (E) inter-subject covariance of the trajectory transition time Σ_τ. For

Table 2
Mean (standard deviation of the mean in brackets) Kendall's tau distance between ground truth event sequence and event sequence returned by the EBM for synthetic data with varying: (A) noise levels, (B) trajectory parameters, (C) time sampling, (D) subtypes. In all tables each column is a new parameter value, and 'estimated' and 'fixed' refer to whether the event distributions are estimated by the EBM or fixed to known values. Bold values represent the default parameters of the EBM. In part (A), experiments A–E represent varying: A. measurement noise; B. inter-subject covariance of trajectory minimum; C. inter-subject covariance of trajectory range; D. inter-subject covariance of trajectory centre; E. inter-subject covariance of trajectory transition time. In part (B), experiments A–B represent varying: A. trajectory centre points, B. trajectory transition times.

A. Noise levels

		0.01	0.1	0.5	1	2
Estimated	A	0.10 (0.06)	0.09 (0.05)	0.11 (0.06)	**0.10 (0.05)**	0.12 (0.05)
	B	0.12 (0.06)	0.00 (0.01)	0.02 (0.02)	**0.09 (0.05)**	0.36 (0.16)
	C	**0.11 (0.05)**	0.11 (0.06)	0.13 (0.06)	0.20 (0.17)	0.18 (0.17)
	D	**0.09 (0.05)**	0.13 (0.04)	0.24 (0.17)	0.32 (0.12)	0.47 (0.18)
	E	**0.10 (0.05)**	0.09 (0.04)	0.15 (0.10)	0.26 (0.12)	0.40 (0.11)
Fixed	A	0.00 (0.01)	0.00 (0.01)	0.01 (0.02)	**0.01 (0.01)**	0.01 (0.02)
	B	0.17 (0.04)	0 (0)	0.00 (0.01)	**0.00 (0.01)**	0.07 (0.04)
	C	**0.00 (0.01)**	0.01 (0.01)	0.03 (0.04)	0.11 (0.04)	0.17 (0.10)
	D	**0.00 (0.01)**	0.01 (0.01)	0.05 (0.04)	0.09 (0.05)	0.17 (0.05)
	E	**0.01 (0.02)**	0.01 (0.01)	0.01 (0.02)	0.10 (0.06)	0.19 (0.06)

B. Trajectory parameters

		1	5	10	15	20
Estimated	A	0.44 (0.21)	0.19 (0.04)	**0.09 (0.04)**	0.08 (0.15)	0.20 (0.22)
	B	0.03 (0.02)	**0.11 (0.05)**	0.19 (0.04)	0.24 (0.05)	0.30 (0.12)
Fixed	A	0.38 (0.22)	0.09 (0.04)	**0.01 (0.01)**	0 (0)	0.01 (0.01)
	B	0.00 (0.01)	**0.00 (0.01)**	0.01 (0.03)	0.04 (0.04)	0.05 (0.03)

C. Time sampling

	0	1	2.5	5
Estimated	0.12 (0.06)	0.10 (0.06)	0.11 (0.07)	**0.10 (0.04)**
Fixed	0 (0)	0.00 (0.01)	0 (0)	**0.00 (0.01)**

D. Subtypes

	0	0.25	0.5	0.75	1
Estimated	**0.10 (0.04)**	0.17 (0.20)	0.50 (0.36)	0.91 (0.06)	0.96 (0.05)
Fixed	**0.01 (0.01)**	0.04 (0.04)	0.45 (0.39)	0.95 (0.04)	1.00 (0.01)

each of the respective simulations we vary the noise level as a proportion p of (A) the estimated measurement noise from ADNI (see Section 2.2.3 and Table 1), and as a proportion p^2 of the covariance matrices: (B) the estimated inter-subject covariance of the trajectory minimum from ADNI (see Section 2.2.3 and Table 1), (C) the square of the mean of the trajectory range μ_r^2 estimated from ADNI (see Section 2.2.3 and Table 1), (D) the range of the trajectory centre points squared, 10^2 years, (E) the range of the baseline time points squared, 20^2 years. For simulations (C), (D), and (E) we assume a diagonal covariance matrix. Varying the measurement noise υ (Table 2A: Experiment A) has little effect on the Kendall's tau distance between the sample event sequences and the ground truth as the estimated measurement noise level is small compared to the inter-subject covariance of the trajectory minimum (Table 1). Varying the inter-subject covariance of the trajectory minimum Σ_a (Table 2A: Experiment B) has a large effect on the Kendall's tau distance for the estimated event distributions, but little effect for fixed event distributions. This shows that it is difficult to estimate the parameters of the event distributions for high biomarker inter-subject variance levels. For very low variance levels on Σ_a the Kendall's tau distance increases again for both fixed and estimated event distributions. This is probably because at very low variance the event distributions do not model the biomarker values over the central portion of the biomarker trajectory, where the trajectory transitions from the minimum to the maximum value. This makes it ambiguous as to whether the biomarker is normal or abnormal during the trajectory transition, making it difficult for the EBM to order the biomarkers. The EBM is robust to inter-subject variation

in the trajectory range Σ_r (Table 2A: Experiment C), giving a similar Kendall's tau distance to the default settings for noise levels up to 50% of the range μ_r. The EBM is quite robust to variation in the trajectory centre points $\Sigma_{c,s}$ (Table 2A: Experiment D) and transition time Σ_τ (Table 2A: Experiment E).

3.1.3. Experiment 2: Trajectory parameters

We performed two experiments to test the robustness of the EBM to different values of the trajectory parameters: (A) varying the centre points of the trajectories, μ_c, and (B) varying the transition time of the trajectories μ_τ. In (A) we assume an evenly spaced set of trajectory centre points over a segment of the disease time course. We vary the duration of this segment as a fraction of 10 years. The trajectory centre points are centred about the middle point along the disease time course (10 years). In (B) we vary the transition time as a fraction of the overall range of the disease time course (20 years), keeping the transition time the same for all biomarkers. The EBM has difficulty estimating the event sequence for both fixed and estimated event distributions when the trajectory centres are close together (Table 2B: Experiment A), and for longer transition times (Table 2B: Experiment B), which violate the assumption of the EBM that an event has either occurred or not occurred. For estimated event distributions the EBM also has difficulty ordering the events when the trajectory centres are spread over the full disease time course (Table 2B: Experiment A). This is because the portion of the trajectory where the biomarker is normal (for early biomarkers) or abnormal (for late biomarkers) is not observed. As in the previous experiments,

fixing the event distributions improves the estimation of the event sequence.

3.1.4. Experiment 3: Time sampling

In this experiment we look at how under-sampling of the disease time course affects the ability of the EBM to recover the sequence of biomarker abnormality. We assume that the time points are sampled from a mixture of three Gaussian distributions, with means at 5 years, 10 years and 15 years respectively. We vary the standard deviation of these distributions, assuming that all of the Gaussians have the same standard deviation. This allows us to simulate the case that cognitively normal, mild cognitive impairment and Alzheimer's disease subjects are at entirely different points along the disease time course. The EBM is robust to under-sampling of the disease time course (Table 2C), giving a similar Kendall's tau distance to the default settings for all simulations.

3.1.5. Experiment 4: Subtypes

To explore the effect of including a set of subjects that follow a different event sequence we modelled two disease subtypes, varying the fraction of subjects that belong to each subtype. For both subtypes the trajectory centre points are evenly spaced from a minimum of 5 years to a maximum of 15 years. In subtype 1 the biomarkers become abnormal in the same order as the default settings: Abeta, p-tau, t-tau, FDG-PET, hippocampal volume, MMSE, ventricles, whole brain volume. In subtype 2 the biomarkers become abnormal in the reverse sequence. This sequence has a Kendall's tau distance of 1 from the sequence of subtype 1. The EBM is robust up to a proportion of 25% of subjects that follow an alternative event sequence (Table 2D): at 25% outliers the Kendall's tau distance is similar to the result for 0% outliers for estimated event distributions, and only slightly increased for fixed event distributions. Likewise, at 75% outliers, when the majority of subjects are subtype 2 the Kendall's tau distance is only slightly worse than for 100% outliers. At 50% outliers the EBM alternates between estimating a sequence similar to subtype 1 and subtype 2.

3.2. DEM stability analysis

We performed a stability analysis of the DEM to test how robust the model is to varying the noise levels on the trajectory parameters.

3.2.1. Default parameter values

Fitting a DEM to each biomarker in turn (see Fig. 3 for estimated synthetic trajectories for FDG-PET) using the default parameter values for the EBM gives an average across biomarkers (excluding CSF for which we only modelled baseline collection) of a mean sample transition time 12.8 ± 1.1 years and standard deviation of this sample transition time of 2.2 ± 1.2 years. This is more than double the simulated trajectory transition time of 5 years. For biomarkers with more data points available (higher biomarker collection rate, e.g. MMSE) the standard deviation of the transition time reduces but the mean transition time remains similar, i.e. the DEM becomes more confident in the biased estimate of the trajectory transition time. In subsequent experiments we simplify the default parameter settings for the DEM to a zero noise case (see Section 2.2.3) to allow us to characterise the types of parameter noise that the DEM is most sensitive to.

3.2.2. Stability of the DEM to noise

For the DEM default parameter values (zero noise case) the DEM is able to recover the trajectory transition time much more accurately (sample transition time is 5.1 ± 0.0 years for a simulated trajectory with a transition time of 5 years). As with the EBM, we ran five simulations to look at the effect of different levels of: (A) measurement noise υ, (B) inter-subject variance of the trajectory minimum Σ_a, (C) inter-subject variance of the trajectory range Σ_r, (D) inter-subject variance of the trajectory centre $\Sigma_{c,s}$, (E) inter-subject variance of the
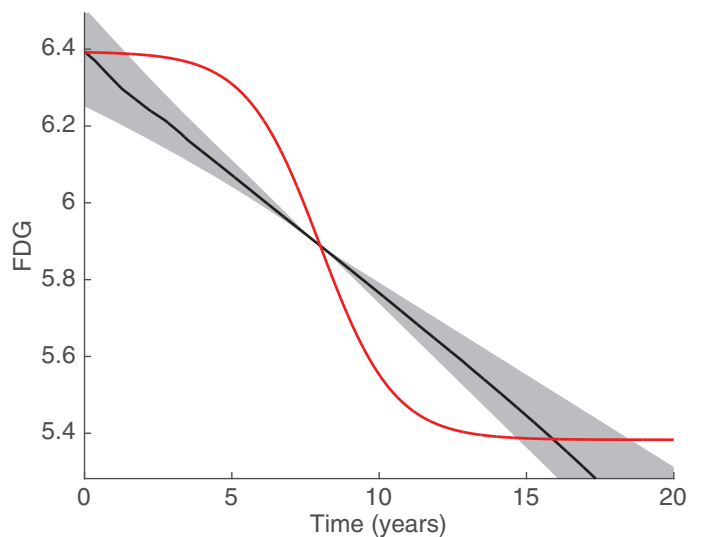
Fig. 3. Integrated DEM trajectories for FDG-PET uptake generated using synthetic data with the default settings for the EBM (Table 1). The ground truth trajectory is in red, and the median estimated trajectory is in black with the inter-quartile range shaded in grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Mean (standard deviation of the mean in brackets) estimated transition time (years) for DEM fitted to synthetic data with varying: A. measurement noise; B. inter-subject variance of trajectory minimum; C. inter-subject variance of trajectory range; D. inter-subject variance of trajectory centre; E. inter-subject variance of trajectory transition time. For all simulations the ground truth transition time is 5 years. Each column is a new parameter value.

	0.1	0.25	0.5	1
A	5.9 (0.4)	9.8 (3.1)	12.7 (4.7)	15.0 (5.2)
B	7.5 (0.2)	11.4 (0.5)	13.3 (0.5)	14.4 (0.7)
C	6.6 (0.2)	9.8 (0.5)	11.6 (0.7)	12.3 (1.0)
D	5.1 (0.0)	5.1 (0.0)	5.1 (0.0)	5.1 (0.0)
E	5.1 (0.0)	5.1 (0.1)	5.4 (0.3)	8.7 (0.9)

trajectory transition time Σ_τ. We vary the noise level as a proportion p of (A) μ_r, and a proportion p^2 of the variance: (B) μ_r^2, (C) μ_r^2, (D) μ_τ^2, (E) μ_τ^2. The DEM is sensitive to measurement noise (Table 3A: Experiment A), and variance of the trajectory minimum (Table 3A: Experiment B) and range (Table 3A: Experiment C), with 25% measurement noise giving a transition time of around twice as long as the actual transition time. The DEM is less sensitive to variance of the trajectory transition time (Table 3A: Experiment E). The DEM is unaffected by noise in the trajectory centre point (Table 3A: Experiment D), as this is removed by differentiating.

4. Discussion

We have presented a framework for the simulation of sporadic neurodegenerative disease datasets. We applied the framework to generate synthetic Alzheimer's disease data, and thereby provide insight into the robustness of the EBM and a DEM to the likely variation in sporadic disease datasets.

4.1. Simulation framework

The simulation framework we have presented is simple and flexible. For example, it is easily extendible to include subjects with a range of demographics, for example age, gender, and education, or genetic risk factors. Such effects can be modelled as a transformation of the trajectory parameters. Here we simplify the diagnosis as a relationship with the biomarker values, however, a more realistic

diagnosis procedure could be simulated that is based on, for example, cognitive test results. It is also possible to add in a screening procedure that post-selects subjects with a similar set of demographics to the dataset being simulated, for example age matching across diagnostic categories. Here we only consider inter-subject variance, however, intra-subject variance for longitudinal datasets could also be modelled.

4.2. Stability analysis

4.2.1. EBM stability analysis

The EBM stability analysis shows that the EBM is sensitive to the estimation of the event distribution parameters. However, when the event distribution parameters are estimated accurately, the EBM is very robust to the likely heterogeneity in sporadic disease datasets. We find that the EBM is robust to noise in the trajectory parameters, different choices of trajectory parameters, under-sampling of the underlying disease time course, and outliers who follow different event sequences.

4.2.3. DEM stability analysis

For all simulations the DEM underestimates the trajectory gradient leading to an over estimation of the trajectory transition time. For the level of noise estimated from the ADNI data this over estimate is more than twice as long as the ground truth trajectory transition time. Whilst this result may be in part due to an over estimation of the amount of inter-subject variation from ADNI, the stability analysis of the DEM shows the DEM will severely over-estimate the trajectory transition time even when the inter-subject variation is much lower. The DEM is very sensitive to measurement noise and inter-subject variation of the trajectory minimum (normal biomarker level) and range (difference between a normal and abnormal level). We further find that using three time points to fit the DEM rather than all available time points, for which the approximation to the derivative is less valid, gives a better estimate of the trajectory transition time, even under high noise levels.

4.2.4. Limitations

In the set of experiments presented we vary each parameter in turn. However, there will likely be multiplicative effects of varying these parameters in combination. We further make a set of assumptions that are specific to hypothetical models of AD, such as sigmoidal trajectories, and to the design of the ADNI dataset, such as the proportion of subjects that drop out per year, and the proportion of subjects in which each biomarker is collected. Other models for the dynamics of the biomarker trajectories, such as statistical time-series models, are easy to incorporate into the simulation framework and would be interesting to consider in future work. We also assume that measurement errors are Gaussian, which may not be the best choice of noise distribution for all of the biomarkers. Therefore, although these simulations do provide an insight into the types of effects that can be expected from different datasets, the simulations should be re-run with dataset specific parameters to assess the performance of the EBM and DEM on alternative datasets. We chose the DEM to be similar to (Villemagne et al., 2013), however other DEM approaches (e.g. Oxtoby et al., 2014) may recover a more accurate estimate of the trajectory transition time for heterogeneous data sets, and should be tested in future work.

4.3. Implications for the application and development of data-driven models

4.3.1. The Event Based Model

The results of the EBM stability analysis show that the EBM is sensitive to the accuracy of the estimated biomarker distribution models, $P(x|E)$, and $P(x|\neg E)$. Therefore the application of the EBM is most effective when the biomarkers have distinct control and case distributions. The results further show that the EBM is robust to 25% outlier corruption, which is higher than the proportion of misdiagnoses we expect in typical sporadic neurodegenerative disease cohorts. The simulations highlight several key areas for improvement of the EBM. First, better estimation techniques for the biomarker distribution parameters should aid recovery of the event sequence when the control and case distributions are not well defined. Adaptation of the EBM to take into account the uncertainty in the biomarker distribution parameters, e.g. by sampling the distribution parameters simultaneously with the ordering, may also help to ameliorate this problem. Second, although the EBM can estimate the ML event sequence for a modest proportion of outliers, it is unable to distinguish other likely event sequences in the data. Future work will look at fitting mixture models with multiple event sequence modes to the data (Young et al., 2015).

4.3.2. Differential equation models

The simulations show that the DEM is sensitive to noise, leading to over estimation of the trajectory transition time, meaning that the DEM should only be applied to biomarkers with low measurement noise and inter-subject variance. Alternatively, robust fitting techniques need to be developed that can correct for the bias encountered when fitting a DEM to noisy biomarker trajectories. The simulations further show that it is important that the duration of follow up for each individual is a good approximation to the derivative (short with respect to the full disease time course). This is shown to be more important than the inclusion of lots of follow up time points, which improves the accuracy of the estimated derivative, suggesting that follow up data over a longer time period should be discarded when fitting a DEM.

5. Conclusion

We have presented a framework for generating synthetic neurodegenerative disease datasets, which can be used to evaluate the robustness of data-driven models to likely variations in sporadic disease datasets, and to directly compare them. We have demonstrated the use of this framework to evaluate the stability of the EBM and a DEM of disease progression to heterogeneity in the ADNI dataset. Future work will use the simulation framework to evaluate the stability of other data-driven models, such as self-modelling regression approaches (Donohue et al., 2014). The simulation framework can further be used as a technique for validating extensions to data-driven models, to determine model weaknesses, and to highlight areas for improvement and future work.

References

Aisen, P.S., Petersen, R.C., Donohue, M.C., Gamst, A., Raman, R., Thomas, R.G., Walter, S., Trojanowski, J.Q., Shaw, L.M., Beckett, L.a, Jack, C.R., Jagust, W., Toga, A.W., Saykin, A.J., Morris, J.C., Green, R.C., Weiner, M.W., 2010. Clinical core of the Alzheimer's disease neuroimaging initiative: progress and plans. Alzheimers. Dement. 6, 239–246. doi:10.1016/j.jalz.2010.03.006.

Bateman, R.J., Xiong, C., Benzinger, T.L.S., Fagan, A.M., Goate, A., Fox, N.C., Marcus, D.S., Cairns, N.J., Xie, X., Blazey, T.M., Holtzman, D.M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P.S., Ghetti, B., Klunk, W.E., McDade, E., Martins, R.N., Masters, C.L., Mayeux, R., Ringman, J.M., Rossor, M.N., Schofield, P.R., Sperling, R.a, Salloway, S., Morris, J.C., 2012. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. N. Engl. J. Med. 367, 795–804. doi:10.1056/NEJMoa1202753.

Bhattacharyya, A.K., 1943. On a measure of divergence between two statistical populations defined by their probability distributions. Bull. Calcutta Math. Soc. 35, 99–109.

Blennow, K., Hampel, H., 2003. CSF markers for incipient Alzheimer's disease. Lancet Neurol. 2, 605–613.

Caroli, a., Frisoni, G.B, 2010. The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. Neurobiol. Aging 31, 1263–1274. doi:10.1016/j.neurobiolaging.2010.04.024.

Clark, C.M., Schneider, J.a., Bedell, B.J., Beach, T.G., Bilker, W.B., Mintun, M.a., Pontecorvo, M.J., Hefti, F., Carpenter, A.P., Flitter, M.L., Krautkramer, M.J., Kung, H.F., Coleman, R.E., Doraiswamy, P.M., Fleisher, A.S., Sabbagh, M.N., Sadowsky, C.H., Reiman, E.P., Reiman, P.E.M., Zehntner, S.P., Skovronsky, D.M., 2011. Use of florbetapir-PET for imaging beta-amyloid pathology. JAMA 305, 275–283. doi:10.1001/jama.2010.2008.

Donohue, M.C., Jacqmin-Gadda, H., Le Goff, M., 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. Alzheimers. Dement. 1–11. doi:10.1016/j.jalz.2013.10.003.

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. doi:10.1016/j.neuroimage.2012.01.062.

Förster, S., Grimmer, T., Miederer, I., Henriksen, G., Yousefi, B.H., Graner, P., Wester, H.-J., Förstl, H., Kurz, A., Dickerson, B.C., Bartenstein, P., Drzezga, A., 2012. Regional expansion of hypometabolism in Alzheimer's disease follows amyloid deposition with temporal delay. Biol. Psychiatry 71, 792–797. doi:10.1016/j.biopsych.2011.04.023.

Fox, N.C., Schott, J.M., 2004. Imaging cerebral atrophy: normal ageing to Alzheimer's disease. Lancet 363, 392–394.

Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M., 2010. The clinical use of structural MRI in Alzheimer disease. Nat. Rev. Neurol. 6, 67–77. doi:10.1038/nrneurol.2009.215.

Herholz, K., 2012. Use of FDG PET as an imaging biomarker in clinical trials of Alzheimer's disease. Biomark. Med. 6, 431–439. doi:10.2217/bmm.12.51.

Jack, C.R., Holtzman, D.M., 2013. Biomarker modeling of Alzheimer's disease. Neuron 80, 1347–1358. doi:10.1016/j.neuron.2013.12.003.

Jack, 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., 2013a. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 12, 207–216. doi:10.1016/S1474-4422(12)70291-0.

Jack, C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9, 119–128. doi:10.1016/S1474-4422(09)70299-6.

Jack, C.R., Vemuri, P., Wiste, H.J., Weigand, S.D., Aisen, P.S., Trojanowski, J.Q., Shaw, L.M., Bernstein, M.a, Petersen, R.C., Weiner, M.W., Knopman, D.S., 2011. Evidence for ordering of Alzheimer disease biomarkers. Arch. Neurol. 68, 1526–1535. doi:10.1001/archneurol.2011.183.

Jack, C.R., Vemuri, P., Wiste, H.J., Weigand, S.D., Lesnick, T.G., Lowe, V., Kantarci, K., Bernstein, M.a, Senjem, M.L., Gunter, J.L., Boeve, B.F., Trojanowski, J.Q., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Knopman, D.S., 2012. Shapes of the trajectories of 5 major biomarkers of Alzheimer disease. Arch. Neurol. 69, 856–867. doi:10.1001/archneurol.2011.3405.

Jack, C.R., Wiste, H.J., Lesnick, T.G., Weigand, S.D., Knopman, D.S., Vemuri, P., Pankratz, V.S., Senjem, M.L., Gunter, J.L., Mielke, M.M., Lowe, V.J., Boeve, B.F., Petersen, R.C., 2013b. Brain β-amyloid load approaches a plateau. Neurology 80, 890–896. doi:10.1212/WNL.0b013e3182840bbe.

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. doi:10.1016/j.neuroimage.2012.07.059.

Klunk, W.E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D.P., Bergstro, M., Savitcheva, I., Debnath, M.L., Barletta, J., Price, J.C., Sandell, J., Lopresti, B.J., Wall, A., Koivisto, P., Antoni, G., Mathis, C.A., Långstro, B., 2004. Imaging brain amyloid in Alzheimer 's disease with Pittsburgh compound-B. Ann. Neurol. 55, 306–319.

Landau, S.M., Mintun, M.a., Joshi, A.D., Koeppe, R.a., Petersen, R.C., Aisen, P.S., Weiner, M.W., Jagust, W.J., 2012. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. Ann. Neurol. 72, 578–586. doi:10.1002/ana.23650.

Lo, R.Y., Hubbard, A.E., Shaw, L.M., Trojanowski, J.Q., Petersen, R.C., Aisen, P.S., Weiner, M.W., Jagust, W.J., 2011. Longitudinal change of biomarkers in cognitive decline. Arch. Neurol. 68, 1257–1266. doi:10.1001/archneurol.2011.123.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–944. doi:10.1212/WNL.34.7.939.

Oxtoby, N.P., Young, A.L., Fox, N.C., Daga, P., Cash, D.M., Ourselin, S., Schott, J.M., Alexander, D.C., 2014. Learning imaging biomarker trajectories from noisy Alzheimer's disease data using a Bayesian multilevel model. In: Simpson, I., Arbel, T., Ribbens, A., Cardoso, M.J., Precup, D. (Eds.), Lecture Notes in Computer Science. Springer International Publishing, pp. 85–94.

Rowe, C.C., Ellis, K.a, Rimajova, M., Bourgeat, P., Pike, K.E., Jones, G., Fripp, J., Tochon-Danguy, H., Morandeau, L., O'Keefe, G., Price, R., Raniga, P., Robins, P., Acosta, O., Lenzo, N., Szoeke, C., Salvado, O., Head, R., Martins, R., Masters, C.L., Ames, D., Villemagne, V.L., 2010. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol. Aging 31, 1275–1283. doi:10.1016/j.neurobiolaging.2010.04.007.

Sabuncu, M.R., Desikan, R.S., Sepulcre, J., Yeo, B.T.T., Liu, H., Schmansky, N.J., Reuter, M., Weiner, M.W., Buckner, R.L., Sperling, R.a., Fischl, B., 2011. The dynamics of cortical and hippocampal atrophy in Alzheimer disease. Arch. Neurol. 68, 1040–1048. doi:10.1001/archneurol.2011.167.

Schott, J.M., Bartlett, J.W., Fox, N.C., Barnes, J., 2010. Increased brain atrophy rates in cognitively normal older adults with low cerebrospinal fluid Aβ1-42. Ann. Neurol. 68, 825–834. doi:10.1002/ana.22315.

Schuff, N., Tosun, D., Insel, P.S., Chiang, G.C., Truran, D., Aisen, P.S., Jack, C.R., Weiner, M.W., 2012. Nonlinear time course of brain volume loss in cognitively normal and impaired elders. Neurobiol. Aging 33, 845–855. doi:10.1016/j.neurobiolaging.2010.07.012.

Villemagne, V.L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K.a, Salvado, O., Szoeke, C., Macaulay, S.L., Martins, R., Maruff, P., Ames, D., Rowe, C.C., Masters, C.L., 2013. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol. 12, 357–367. doi:10.1016/S1474-4422(13)70044-9.

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. doi:10.1093/brain/awu176.

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., 2015. Multiple Orderings of Events in Disease Progression. LNCS 9123, 711–722. doi:10.1093/brain/awu176.