

Accepted Manuscript

An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease

Hubert M. Fonteijn, Marc Modat, Matthew J. Clarkson, Josephine Barnes, Manja Lehmann, Nicola Z. Hobbs, Rachael I. Scahill, Sarah J. Tabrizi, Sebastien Ourselin, Nick C. Fox, Daniel C. Alexander

PII: S1053-8119(12)00079-1
DOI: doi: [10.1016/j.neuroimage.2012.01.062](https://doi.org/10.1016/j.neuroimage.2012.01.062)
Reference: YNIMG 9127

To appear in: *NeuroImage*

Received date: 20 July 2011
Revised date: 23 December 2011
Accepted date: 8 January 2012



Please cite this article as: Fonteijn, Hubert M., Modat, Marc, Clarkson, Matthew J., Barnes, Josephine, Lehmann, Manja, Hobbs, Nicola Z., Scahill, Rachael I., Tabrizi, Sarah J., Ourselin, Sebastien, Fox, Nick C., Alexander, Daniel C., An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease, *NeuroImage* (2012), doi: [10.1016/j.neuroimage.2012.01.062](https://doi.org/10.1016/j.neuroimage.2012.01.062)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**An event-based model for disease progression and its application in familial
Alzheimer's disease and Huntington's disease**

Hubert M. Fonteijn^{a,b,c}, Marc Modat^{a,d}, Matthew J. Clarkson^{a,d,e}, Josephine
Barnes^e, Manja Lehmann^e, Nicola Z. Hobbs^f, Rachael I. Scahill^f, Sarah J.
Tabrizi^{f,g}, Sebastien Ourselin^{a,d,e}, Nick C. Fox^{e,g}, and Daniel C. Alexander^{a,b}

^aCentre for Medical Image Computing, University College London, Gower Street,
WC1E 6BT, London, UK

^bDepartment of Computer Science, University College London, Gower Street,
WC1E 6BT, London, UK

^cMax Planck Institute for Psycholinguistics, P.O. Box 9101, 6500 HB, Nijmegen,
The Netherlands

^dDepartment of Medical Physics and Bioengineering, University College London,
Gower Street, WC1E 6BT, London, UK

^eDementia Research Centre, UCL Institute of Neurology, University College
London, 8-11 Queen Square, WC1N 3AR, London, UK

^fDepartment of Neurodegenerative Disease, UCL Institute of Neurology,
University College London, Queen Square, WC1N 3BG, London, UK

^gDepartment of Clinical Neurology, National Hospital for Neurology and
Neurosurgery, Queen Square, WC1N 3BG, London, UK

Corresponding author: Hubert M. Fonteijn

Max Planck Institute for Psycholinguistics, P.O. Box 310

Telephone: +31 24 3668294

Fax: +31 24 3610989

e-mail: Hubert.fonteijn@fcdonders.ru.nl

E-mail addresses other authors:

Marc Modat: m.modat@ucl.ac.uk

Matthew J. Clarkson: m.clarkson@ucl.ac.uk

Josephine Barnes: j.barnes@dementia.ion.ucl.ac.uk

Manja Lehmann: lehmann@drc.ion.ucl.ac.uk

Nicola Z. Hobbs: n.hobbs@ucl.ac.uk

Rachael I. Scahill: r.scahill@ucl.ac.uk

Sarah J. Tabrizi: s.tabrizi@prion.ucl.ac.uk

Sebastien Ourselin: s.ourselin@cs.ucl.ac.uk

Nick C. Fox: n.fox@ucl.ac.uk

Daniel C. Alexander: d.alexander@cs.ucl.ac.uk

Abstract

Understanding the progression of neurological diseases is vital for accurate and early diagnosis and treatment planning. We introduce a new characterization of disease progression, which describes the disease as a series of events, each comprising a significant change in patient state. We provide novel algorithms to learn the event ordering from heterogeneous measurements over a whole patient cohort and demonstrate using combined imaging and clinical data from familial-Alzheimer's and Huntington's disease cohorts. Results provide new detail in the progression pattern of these diseases, while confirming known features, and give unique insight into the variability of progression over the cohort. The key advantage of the new model and algorithms over previous progression models is that they do not require *a priori* division of the patients into clinical stages. The model and its formulation extend naturally to a wide range of other diseases and developmental processes and accommodate cross-sectional and longitudinal input data.

Keywords:

- Disease progression
- MRI
- Alzheimer's disease
- Huntington's disease

1 Introduction

Modeling patterns of disease progression is a key aim of medical science. Such patterns further our understanding of the disease and help construct staging systems that assist diagnosis and treatment. For example, the Ann Arbor system (Carbone et al., 1971) classifies lymphomas into four stages that progress from a localized tumor within a single lymph node to a widespread involvement of lymph nodes and other organs.

Disease progression occurs at various levels, ranging from the symptoms a patient experiences to cellular and biochemical changes. For instance, the first symptoms of Alzheimer's disease (AD) are a loss of episodic memory, followed by a progressive deterioration of other cognitive abilities, such as language and executive function. However, cellular pathology precedes these symptoms, often by many years (Dickerson et al., 2009; Scahill et al., 2002; Thompson et al., 2003; Thompson et al., 2001). This cellular pathology includes amyloid plaques and intracellular neurofibrillary tangles (NFTs), which are linked to neuronal degeneration and loss (Herrup, 2010). The spread of NFTs has a consistent pattern starting in memory-related areas, such as the hippocampus and the entorhinal cortex, progressing to other higher cognitive areas and finally the primary cortices (Braak and Braak, 1991).

The importance of understanding the progression patterns of neurological diseases is reflected in recent efforts to collect large-scale multi-centre data sets combining clinical,

imaging and pathology measurements from hundreds of patients, suffering from Alzheimer's disease (Mueller et al., 2005) or Huntington's disease (Tabrizi et al., 2009). Despite the availability of these data, current models of disease progression (Dickerson et al., 2009; Jack et al., 2010; Scahill et al., 2002; Thompson et al., 2003; Thompson et al., 2001) remain crude. These models use symptomatic staging to divide patients into a small number of groups, e.g. "presymptomatic", "mild", "moderate" or "severe", and assess the differences in biomarkers among those groups. Symptomatic staging is an invaluable tool to explain disease progression to patients and their families. However, it relies on imprecise and subjective clinical assessment, which limits the temporal resolution of progression models constructed this way and thus their power to discriminate diseases and stage patients. The need is therefore urgent for techniques that exploit large-scale heterogeneous data sets to obtain more detailed models of disease progression.

In this paper, we introduce a direct computational approach to modeling disease progression. The new model describes the disease as a sequence of discrete events. In our model, the events are changes in patient state, such as the onset of a new symptom ("patient shows reduced memory performance") or the first measurement of tissue pathology ("lumbar puncture shows reduced beta amyloid"). The computational aim is to find the ordering of the events that is most consistent with a set of measurements from a cohort of patients. Thus the event-based model translates naturally into a semantic description of the disease similar to a clinician's intuitive description of its progression.

We formulate the event-based disease progression model and develop new Bayesian statistical algorithms for fitting that refine an earlier version of this model in (Fonteijn et al., 2011). We demonstrate the model and fitting procedure on data from two cohorts of patients: a familial Alzheimer's disease (fAD) cohort and a Huntington's disease (HD) cohort. FAD is a rare autosomal-dominantly inherited variant of AD, which causes early onset (Bertram et al., 2010). HD is also an autosomal-dominantly inherited disease (The Huntington's Disease Collaborative Research Group, 1993) and is characterized by motor and cognitive impairment and neuropsychiatric disturbance. The disease progression patterns of both fAD (Ridha et al., 2006; Scahill et al., 2002) and HD (Paulsen et al., 2008; Rosas et al., 2008; Tabrizi et al., 2011) have been well studied using clinical staging and imaging, which provides useful validation for the event-based model. The models here consist of two types of events: clinical events and atrophy events. Clinical events are transitions to a later clinical status, such as from presymptomatic AD to Mild Cognitive Impairment (MCI), whereas atrophy events correspond to the occurrence of significant regional atrophy within a patient. However, the formulation extends easily to include other event types.

The remainder of this paper is structured as follows: section 2 formulates the event-based disease progression model and develops algorithms for fitting. Section 3 provides more detail about the fAD and the HD data sets, in terms of demographics and data acquisition procedures, and outlines the preprocessing pipeline that we use. Section 4 summarizes the

results of fitting the event-based disease progression model to both cohorts and further experiments to demonstrate the model's ability to stage unseen data and its consistency between hemispheres. Section 5 discusses the implications of these findings, indicates what the current limitations of the model are and concludes with suggestions for future work.

2 Theory

This section develops the event-based model and the algorithms for estimating the model. Section 2.1 formally introduces the event-based disease progression model and derives a forward model from an event ordering to the set of measurements acquired from the patient cohort. The forward model provides a data likelihood function, given an event sequence. Section 2.2 casts the estimation of a characteristic event sequence for a particular data set into a Bayesian framework and introduces a Markov Chain Monte Carlo algorithm for sampling from the posterior distribution on the event sequence. Section 2.3 details the particular mixture model for the likelihood function of individual measurements required to complete the forward model in section 2.1 for the particular data sets and event types that we use here. Section 2.4 discusses how the model is adapted to include different types of events.

2.1 The event-based disease progression model

Figure 1 illustrates the event-based disease progression model and the algorithms we devise to estimate it. The model consists of a set of events E_1, \dots, E_N and an ordering $S=(s(1), \dots, s(N))$, which is a permutation of the integers $1, \dots, N$ determining the event ordering $E_{s(1)}, \dots, E_{s(N)}$. The set of events is specified *a priori* and we estimate S from a data set X , which contains a set X_j of measurements from each patient $j=1, \dots, J$ and X_l from each control $l=1, \dots, L$. Each set $X_j = \{x_{1j}, x_{2j}, \dots, x_{Nj}\}$, or similarly X_l , contains measurements x_{ij} , that are informative about the occurrence of event E_i in patient j .

Two key assumptions of the event-based model are: first, measurements decrease monotonically as the disease progresses; second, the event sequence is consistent over all patients. The first assumption is in line with existing models of disease progression, such as the hypothetical model for AD in (Jack et al., 2010). Thus, patients for whom event E_i has occurred cannot revert to an earlier state where E_i has not occurred. This assumption is essential because it ensures that snapshots are informative about the full event ordering. The second assumption is essential to enable pooling of snapshots from individual patients to inform the complete event sequence for the whole cohort.

Fitting the event-based disease progression model requires the evaluation of the probability density function $p(S | X)$ of a particular event ordering given the data. We start by fitting simple models for the likelihood function $p(x_{ij} | E_i)$ on the measurement x_{ij} given that E_i has occurred and, similarly, $p(x_{ij} | \neg E_i)$ on x_{ij} given that E_i has not occurred. We assume that both distributions are independent of j . In the remainder of this section

we treat these distributions as known. Section 2.3 explains how we construct models for these likelihoods.

If patient j is at position k in the progression model, events $E_{s(1)}, \dots, E_{s(k)}$ have occurred, while events $E_{s(k+1)}, \dots, E_{s(N)}$ have not, and we can write the likelihood of that patient's data given S as

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

where we assume that individual measurements are independent. We integrate out the hidden variable k to obtain

$$p(X_j | S) = \sum_{k=0}^N p(k) p(X_j | S, k), \quad (2)$$

where $p(k)$ is the prior probability of being at position k in the ordering; we assume a uniform prior on k here. Next, we assume independence of measurements from different patients to obtain

$$p(X | S) = \prod_{j=1}^J p(X_j | S). \quad (3)$$

Combining Eqs. 1, 2 and 3 yields the total likelihood

$$p(X | S) = \prod_{j=1}^J \left[\sum_{k=0}^N p(k) \left(\prod_{i=1}^k p(x_{s(i),j} | E_{s(i)}) \prod_{i=k+1}^N p(x_{s(i),j} | \neg E_{s(i)}) \right) \right]. \quad (4)$$

2.2 Model estimation

We use Bayes' theorem to obtain the posterior distribution

$$p(S | X) = \frac{p(S)p(X | S)}{p(X)}. \quad (5)$$

The marginal distribution $p(X)$ is analytically intractable. We therefore use a MCMC algorithm (Gilks et al., 1996) to sample from $p(S | X)$. We use flat priors on the ordering S , on the assumption that *a priori* all orderings are equally likely.

The MCMC algorithm proceeds as follows: at each iteration t , a perturbation S' of the current model S_t swaps the positions of two randomly chosen events. Next, $S_{t+1} = S'$ with probability $\min(a, 1)$ where $a = p(X | S')/p(X | S_t)$ is the likelihood ratio; otherwise $S_{t+1} = S_t$. MCMC algorithms depend on accurate initialization to achieve good mixing properties within realistic time spans. We therefore initialize the MCMC algorithm with an initial sequence close to the maximum likelihood solution S_{ML} for $p(X | S)$. We find the initial sequence with a greedy ascent algorithm from a random starting point. The greedy ascent algorithm uses the same perturbation rule as the MCMC algorithm, but sets $a=1$ for $p(X | S') > p(X | S_t)$ or $a=0$ otherwise. We run the greedy algorithm for 2,000 iterations, which generally is enough to reach a local maximum. We repeat the greedy algorithm 10 times from different initialization points, which provides high likelihood of reaching the global maximum in at least one iteration. Next, we run the MCMC procedure for 1,100,000 iterations, of which we discard 100,000 burnin iterations and retain the remaining 1,000,000 samples.

We derive a characteristic event ordering and a positional variance diagram expressing the uncertainty in the ordering from the set of 1,000,000 MCMC samples (see figure 1D). We compute the characteristic ordering \bar{S} , by ordering the events by their average position in the MCMC samples. We suggest this characteristic ordering, rather than for example simply the maximum likelihood ordering, because it contains additional information about the separation of consecutive events in the characteristic ordering that may be useful, although we do not exploit this extra information here. The positional variance diagram illustrates the posterior distribution on S and shows a histogram of the position in the ordering of each event over the set of MCMC samples.

2.3 Mixture models for the data likelihood

Evaluating equation 4 requires separate models for both $p(x_{ij} | E_i)$ and $p(x_{ij} | \neg E_i)$, which we estimate from the set of measurements $Y_i = \{x_{ij} | j=1, \dots, J\} \cup \{x_{il} | l=1, \dots, L\}$ associated with event E_i from all the controls and patients. However, the fact that a subject is a patient or a control does not determine whether E_i has occurred or not within that subject. This necessitates fitting a mixture model to the data, whose components correspond to $p(x_{ij} | E_i)$ and $p(x_{ij} | \neg E_i)$ respectively.

The particular choice of model for each component depends on the type of event and the measurement that informs on its occurrence. Here, the measurements we use are predominantly measurements of grey matter atrophy obtained from comparison of anatomical MR images to an earlier baseline. The atrophy measurement is always

positive, unity corresponds to no volume change and values less than one indicate atrophy. Therefore we fit a mixture of a Gaussian and uniform distribution in which the Gaussian component corresponds to $p(x_{ij} | \neg E_i)$, while the uniform distribution corresponds to $p(x_{ij} | E_i)$. The distributions of atrophy measurements from controls appear single modal for all regions in our data sets and the Gaussian model thus provides a simple and effective model. We use a uniform distribution on $[m_1, m_2]$ for $p(x_{ij} | E_i)$ to reflect that the measurement gradually decreases as the severity of the event increases and patients may have any level of severity.

Our limited data sets do not provide good coverage of the whole range of severity, as the cohorts are skewed towards early disease stages, so we impose the uniform shape. The lower bound m_1 is smaller than any atrophy measurement we observe, whereas the upper bound m_2 is set to the mean of the Gaussian component that represents individuals in which E_i has not occurred. Under this model, subjects with severe atrophy have the same $p(x_{ij} | E_i)$ as subjects whose atrophy is comparable to controls. This might seem counterintuitive, as perhaps $p(x_{ij} | E_i)$ should be larger for more severe atrophy. However, the important quantity in the MCMC sampling is the ratio $p(x_{ij} | E_i) / p(x_{ij} | \neg E_i)$, which is much larger for severe atrophy, because $p(x_{ij} | \neg E_i)$ is smaller.

In the fitting procedure, we assume that the controls contain no subjects in which E_i has occurred, while the patients contain both classes of subjects. We therefore start by fitting the Gaussian component to the data from the controls separately and then fit the mixture

to the complete data set, while keeping the Gaussian component's parameters fixed, using the Expectation Maximization algorithm (Dempster et al., 1977).

Earlier work (Fontejn et al., 2011) used a Gaussian mixture model for $p(x_{ij} | E_i)$ instead of the uniform model we use here. The two models in fact produce similar results for the fAD cohort, which contains patients at a range of different disease stages. However, the uniform model behaves much better for the HD cohort. The HD cohort is heavily skewed towards earlier stages, so, in each region, less patients tend to show atrophy than in the fAD cohort. In many regions, the distribution of atrophy values over the whole cohort appears unimodal, because only a few patients show atrophy. This impedes separating the data into two Gaussian components. The uniform distribution, on the other hand, guarantees a long tail into the abnormal part of atrophy distribution for each region and thus better separates $p(x_{ij} | E_i)$ and $p(x_{ij} | \neg E_i)$.

2.4 Modeling different event types

The event-based disease progression model of the fAD cohort includes both clinical events and atrophy events. The atrophy events can be treated exactly as in section 2.3, but the clinical events require slightly different treatment, because they are binary rather than informed by a measurement: either the patient was classified as AD from the cognitive tests or they were not. We simply assign fixed high and low likelihood values to $p(x_{ij} | E_i)$ and $p(x_{ij} | \neg E_i)$, respectively if the event has occurred, or vice versa if not. These likelihood values were selected to be representative for the range of values we observe

for the atrophy events. The values we use in fact come from fitting a mixture of two Gaussians to $p(x_{ij} | \neg E_i)$ over all *atrophy* events and subjects. The means of both components provide the fixed high and low likelihood values of $p(x_{ij} | \neg E_i)$ for the *clinical* events. We repeat this procedure for $p(x_{ij} | E_i)$. However, in practice, the ordering of the clinical events A, B and C (which correspond to changes in clinical diagnosis, see section 3.1) does not critically depend on the exact choice for these likelihood values: using likelihood values that are randomly drawn from the mixture distributions of likelihood values does not shift the position of events A and B, whereas the shift in the position of event C is limited to maximally 2 places in more than 95 % of the cases (see figure 3 for the positions of events A, B and C when using the mixture of Gaussians to determine the fixed high and low likelihood values). Moreover, changing the likelihood values of the clinical events always leaves the ordering of the atrophy events unaffected.

3 Materials and methods

We demonstrate the event-based disease progression model on data from two cohorts. Both data sets are longitudinal and contain multiple measurements per patient. However, for the purposes of demonstrating the model here, we treat each snapshot as independent as if it came from a separate patient. We use two types of event: clinical events and atrophy events. Clinical events are transitions to a later clinical status, such as from presymptomatic AD to Mild Cognitive Impairment (MCI). Criteria for transition are outlined in section 3.1. Atrophy events correspond to the occurrence of significant

regional atrophy within a patient. We compute regional atrophy values from serial T1-weighted MRI data, about which we provide more details in section 3.2. Section 3.3 outlines the preprocessing steps which are used to compute regional atrophy values.

3.1 Subjects

The fAD data set is the same as used in (Ridha et al., 2006). Briefly, nine carriers of autosomal mutations associated to AD were recruited from the Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery. 25 Age-matched and sex-matched controls (two to three controls per mutation carrier) were also recruited from spouses, relatives and healthy volunteers. All participants gave written informed consent as approved by the local ethics committee. All mutation carriers underwent comprehensive clinical and neuropsychological assessments, including the mini-mental state examination (MMSE) (Folstein et al., 1975) and volumetric MRI scans, at each visit (41 visits: three to eight per patient, time interval between visits \approx 1 year). All controls underwent the same procedure (54 visits, two for each participant, except for two participants who had four scans each). The clinical status of each mutation carrier was classified at each time point as: 1) presymptomatic, if participants fell short of both NINCDS-ADRDA and MCI criteria; 2) MCI, if the patient fulfilled MCI criteria (Petersen et al., 1999); and 3) familial AD, if the patient fulfilled National Institute of Neurological and Communicative Disorders and Stroke and AD and Related Disorders (NINCDS-ADRDA) diagnostic criteria for probable AD (McKhann et al., 1984). The transitions between these clinical statuses constitute the clinical events: event A indicates

that the patient is presymptomatic, event B that the patient has progressed from presymptomatic to MCI and event C indicates the transition from MCI to familial AD. All mutation carriers became affected during the follow-up period and so have a known date of onset, except one who has remained presymptomatic.

The HD data have previously been analyzed in Henley et al. (Henley et al., 2009). 61 patients with genetically-confirmed HD were recruited from the multidisciplinary HD clinic at the National Hospital for Neurology and Neurosurgery, London and the Huntington's Disease Clinic at Addenbrooke's Hospital, Cambridge. 40 patients were classified as "early HD" (stages 1 and 2) (Shoulson and Fahn, 1979), and 21 patients were gene carriers without motor signs, i.e. premanifest. Twenty neurologically normal controls, either patients' partners or at-risk subjects who had tested negative for the HD gene expansion, were also recruited. All subjects gave written informed consent in accordance with the Declaration of Helsinki, and the study had local research ethics committee and Hospital Trust approval. Subjects underwent assessment by a neurologist and psychologist, using the UHDRS clinical rating scale (Huntington's Study Group, 1996) which comprised motor, functional and cognitive assessments. Subjects were scanned at baseline and after 12 months. A subset of the subjects (13 controls, 16 premanifest patients and 25 early HD patients) were also scanned at 24 months after baseline.

3.2 Image acquisition

Imaging on the fAD cohort was undertaken with two scanners: in the period between 1991 and 2000 a 1.5 T GE Signa MRI scanner (General Electric Medical Systems, Waukesha, WI, USA) was used. T1 images were acquired using a Spoiled Gradient Echo technique with the following parameters: (256 x 256 matrix, field of view 24 x 24 cm, in-plane resolution 0.9375 x 0.9375 mm, TR/TE/NEX/FA=35 ms/5 ms/1/35°) yielding 124 contiguous 1.5 mm thick slices. Between 2000 and 2005 scans were acquired on a different 1.5 T GE Signa scanner. On this scanner T1 images were acquired using an inversion recovery fast spoiled gradient echo sequence with the following parameters: 256 x 256 matrix, field of view 24 x 18 cm, in-plane resolution 0.9375 x 0.9375 mm, TR/TE/TI/NEX/FA=14 ms/5.4 ms/650 ms/1/15°, yielding 124 contiguous 1.5 mm thick slices.

Imaging for the HD data set was undertaken using a 1.5 T GE Signa MRI scanner. T1 images were acquired using an inversion recovery prepared fast spoiled gradient echo sequence with the following parameters: 256 x 256 matrix, field of view 24 x 18 cm, in-plane resolution 0.9375 x 0.9375 mm, TR/TE/TI/NEX/FA = 13 ms/5.2 ms/650 ms/1/13°, yielding 124 contiguous 1.5 mm thick coronal slices.

3.3 Data preprocessing

The preprocessing pipeline extracts anatomically defined regions, determines voxel-wise atrophy using non-rigid registration methods and combines both steps to compute regional atrophy values.

We use FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>) (version 4.5) to segment the brain into cortical and subcortical structures. FreeSurfer starts with gray matter/white matter segmentation. It then uses a labeling procedure on each subject's data which is based on a training set of manually segmented scans (Fischl et al., 2002; Fischl et al., 2004). Thus an anatomical segmentation of cortical and subcortical gray matter is achieved which is dependent on each subject's individual anatomy and which has similar accuracy to manual labeling (Fischl et al., 2004). In the fAD data, we supplement FreeSurfer's segmentation with subject-specific hippocampus masks, which were manually traced by Ridha et al. (Ridha et al., 2006).

We calculate voxel-wise atrophy values using non-linear registration methods. We use a free form deformation registration method (Modat et al., 2010), which first uses a rigid registration to realign the repeat scans to the baseline scan and then uses a set of cubic B-splines to locally deform the repeat scan to match the baseline scan. In each voxel, the change that is required by this matching procedure is quantified by the determinant of the Jacobian of the deformation field. We calculate regional atrophy by calculating the median Jacobian determinant for each segmented region.

4 Results

This section starts by demonstrating the convergence and mixing properties of the algorithm (section 4.1), before describing the characteristic event sequences for the fAD and HD cohorts (sections 4.2 and 4.3 respectively). Sections 4.4 and 4.5 then discuss the

positional variance diagrams for both cohorts. The last two sections describe additional experiments that demonstrate the robustness and validity of the event-based disease progression model.

4.1 Convergence and mixing properties

Figure 2a shows traces from the greedy ascent algorithm that is used to initialize the MCMC algorithm in the fAD cohort. The traces are similar for the HD cohort and are therefore not shown. Although the ten iterations are randomly initialized, they converge to very similar solutions within 2000 iterations. The log likelihood is several orders of magnitude higher for the fitted model than random starting points suggesting that the maximum likelihood model is well defined. Figure 2b shows the MCMC trace, which suggests that the MCMC chain indeed converges when initialized with the greedy ascent algorithm and with the current burnin period.

4.2 Event sequence of fAD cohort

Figure 3 shows the characteristic event sequence \bar{S} for the fAD cohort. We average regional atrophy across hemispheres to reduce the number of events and the model's complexity (section 4.7 shows experiments that include regions from both hemispheres separately). Significant atrophy first occurs in the hippocampus, the precuneus and the inferior parietal cortex, followed by temporal neocortical regions. Other parietal and frontal areas become involved subsequently, while primary cortices are only affected in late stages. This progression pattern broadly agrees with how neurofibrillary tangles

spread through the brain (Braak and Braak, 1991). An interesting difference in our findings is the relatively late involvement of the entorhinal cortex, which is known to show early pathology, including atrophy (Cardenas et al., 2003; Du et al., 2004; Xu et al., 2000). However, the small size of the entorhinal cortex makes it difficult to segment automatically and therefore increases the variance of the automatically generated atrophy measures. Significance of entorhinal atrophy is thus harder to establish than in larger regions, which shifts the event later in \bar{S} . Manual tracing of the entorhinal cortex (Cardenas et al., 2003; Du et al., 2004; Xu et al., 2000) would likely move the event sooner. The positions of clinical events B (MCI diagnosis) and C (AD diagnosis) reveal new information that previous models cannot: the set of regions that, on average, show significant atrophy at the time of diagnosis. For example, only the hippocampus shows an atrophy measurement substantially different to controls prior to clinical MCI.

4.3 Event sequence of HD cohort

Figure 4 shows \bar{S} for the HD cohort. The early HD events occur in the putamen, the caudate and the thalamus. Both pathological studies (Vonsattel et al., 1985) and imaging studies (Aylward et al., 2010; Tabrizi et al., 2011) support the early involvement of these regions. Atrophy then spreads towards the cuneus, the superior frontal cortex and the precentral cortex, in agreement with the progression pattern found by Rosas et al. (Rosas et al., 2008) at various clinical stages of HD. The early occurrence of cingulate atrophy is supported by other imaging studies (Henley et al., 2009; Tabrizi et al., 2009) and is in

line with the observed deficits in mood and emotional processing in HD patients (Tabrizi et al., 2009).

4.4 Positional variance for the fAD cohort

Figure 5 shows the positional variance diagram for each cohort. These diagrams show the histogram of positions of each event over the MCMC samples and thus express the posterior distribution on the event sequences. Although the diagrams require some care in interpretation, since they are posterior distributions on a single ordering rather than true distributions of orderings, they provide unique insight into the variance of the progression pattern, which no previous models provide.

In the fAD positional variance diagram, the separation of events that are consecutive in \bar{S} is often weak, but events further apart in \bar{S} show no overlap in positional variance, demonstrating that their temporal separation is consistent over the cohort and a strong feature of the disease progression. Moreover, we observe consecutive clusters of regions, where ordering within the clusters is weak (extensive overlap of region positions within the cluster), but ordering between clusters is strong (little or no overlap between the clusters themselves).

An appealing interpretation of these clusters is that they represent a division of the events into separate stages within which events may occur in any order, but between which events rarely switch position. However, a weak local ordering of events can be caused by a number of factors. First, weakly separated events can arise when the event sequence is

variable over the population, or when these events occur simultaneously, which would justify the interpretation of clusters of events as disease stages. Second, weak local ordering can also arise when the data set does not provide snapshots at all stages, as is likely the case in the fAD data where the number of snapshots (41) is of the same order as the number of events (38). This undersampling also introduces blocks of events whose ordering is undetermined by the data. Thus, although \bar{S} estimated from the fAD cohort may exhibit a degree of overfitting in this way, our formulation of the model and its estimation mitigate the effect, because the positional variance diagrams reveal which parts of the ordering the data support.

4.5 Positional variance for the HD cohort

The HD positional variance diagram shows strong grouping of the early events, but a large block of later events has only weak relative ordering. However, the cohort does not include any late stage patients and therefore evidence for any particular ordering of late events is expected to be weak. Moreover, HD is pathologically and clinically heterogeneous and different clinical presentations are likely to have different patterns of regional atrophy (Rosas et al., 2008). This, as well as the larger number of patients and fewer repeats, is likely to contribute to the greater positional variance than we observe in fAD.

4.6 Staging patients within the event-based model

The event-based disease progression model provides a natural framework for fine-grained staging of patients. We can use \bar{S} to determine the stage k for each patient which maximizes the data likelihood, i.e. the k that maximizes $p(X_j | \bar{S}, k)$ in Eq. 1. In the fAD cohort, there are multiple measurements for each patient. A simple check on the validity of the model investigates whether later time points within each patient correspond to later stages in the disease progression model. To provide a proper test for unseen data, we fit the model separately for each time point, excluding that time point, and estimate the position of the excluded time point in the model of the remaining data. Figure 6 shows indeed that generally later time points within patients correspond to later stages. Exceptions arise for scans acquired during the late stages of the disease. Figure 5 shows that there is high uncertainty around the relative ordering of these late stages, which makes the ordering of the later time points more difficult to discern.

4.7 Consistency of event-based models between hemispheres

The disease progression models in figures 3-5 use regional atrophy values averaged across hemispheres. To test consistency between hemispheres, we fit separate models for atrophy events in left and right hemisphere regions. Figure 7 shows that the ordering is largely consistent across hemispheres for the fAD cohort. The upper right block of events shows little correspondence in event position across hemispheres. However, these events correspond to the lower right block of events in figure 4 in which there is no substantial deviation from controls so the ordering is weak.

5 Discussion

The event-based disease progression model has major advantages over current approaches to modeling the progression of neurological and other diseases. First, it directly extracts the time line of disease progression from the data, instead of relying on clinical staging. Moreover, the model uniquely characterizes uncertainty in the event ordering (figure 5), arising from measurement noise and heterogeneity over the population. Characterizing uncertainty provides new insight and is essential for model-based discrimination of different diseases and diagnosis. Second, model fitting naturally exploits longitudinal or cross-sectional data sets or, as we demonstrate here, heterogeneous data sets with different numbers of time points for each patient. Third, the approach combines information from different sources naturally. Here, for example, we demonstrate orderings of events including regional atrophy events informed by imaging as well as clinical events informed by patients' scores in cognitive tests. Events informed by other types of measurement, such as blood tests, other imaging modalities, lumbar puncture, or biopsy, are straightforward to incorporate as new event types. This is a key area for further work, since including different measurement modalities enables events to be informed by multiple measurements, which would relax the one-to-one mapping between measurements and events and lead to more robust event detection.

These advantages make the model and estimation procedure very suitable for application in large multi-modal data sets, such as the ADNI data set for AD (Mueller et al., 2005) and the TrackHD data set for HD (Tabrizi et al., 2009). Such experiments will provide complete progression models for AD, HD and a variety of other diseases, which promise great advances in early and accurate diagnosis. The new method scales up easily to larger data sets, as computation is linear in both the number of patients and events. Some complications are likely to arise from misclassifications, because some large data sets do not have confirmed diagnosis for the whole cohort. For example, clinical diagnosis in sporadic AD is imprecise and large cohorts therefore potentially contain a proportion of misdiagnosed patients and undiagnosed controls. Future work must quantify the robustness of the method to the misclassification inherent in more general large-scale studies.

The event-based model offers a natural framework to compare progression patterns among phenotypes in AD, HD or any other disease. In a similar way, comparison of models of treatment and control groups in clinical trials can reveal the effect of a treatment on the progression, potentially providing new markers of treatment efficacy. Moreover, staging based on the event-based model (see section 4.6) offers promise as a more reliable basis for treatment plans than clinical staging.

Other classes of model are similar to the event-based model. The most comparable model fits time lines of the emergence and extinction of species to the fossil record from

different sites (Puolamäki et al., 2006). The main difference with the event-based model is the fact that species can appear *and disappear* from the fossil record. This makes this model symmetric with respect to time reversal and necessitates the input of prior information to break this symmetry. The event-based model is already asymmetric with respect to time reversal, because we assume that, once a patient shows evidence that an event has occurred, this evidence will remain present at later time points. Another related class of models is state-space models, such as Hidden Markov Models (HMM), which have found widespread use in areas such as speech recognition (Rabiner, 1989) and gene finding (Korf, 2004). Although both state-space models and the event-based model are similar in that they use a set of abstract states (disease stages in the event-based model) to represent patterns of observed data, important differences exist: first, state-space models generally require input data with known temporal ordering, which the data we model does not have. Second, the transition probabilities between states in state-space models are generally learned from the data, whereas the transition probabilities between stages in the event-based model are defined implicitly (1 for subsequent stages, 0 for all other combinations of stages).

The model and estimation procedure in their current form make various assumptions that must be considered when interpreting results. First, we emphasize that the nature of the events in the model is “pathology becomes detectable through a particular measurement” rather than “pathology appears”. The sensitivity of the measurements to pathology limits the ability of any progression model to reflect the underlying order of pathology. To

make this clear, we refer to the idealized ordering of pathology occurrences as the *pathology ordering* to distinguish it from \bar{S} , which is the order in which events become detectable. Differences in the sensitivity of event measurements (heteroscedasticity) cause departures of \bar{S} from the pathology ordering. The relatively late position of the entorhinal cortex we observe in fAD (see figure 3) provides an example: inaccurate segmentation of the entorhinal cortex reduces sensitivity of the atrophy measurement relative to other regions; a significant deviation of atrophy from controls therefore occurs later than in regions that are later in the pathology ordering but whose atrophy measurements are more sensitive. In practice, diagnosis and staging mechanisms must rely on measurements. Therefore, the model of detectability through measurement, which \bar{S} provides, has direct practical utility, but we must be cautious in inferring the idealized pathology ordering.

Second, equation 1 assumes independence of the likelihood of each measurement conditional on event occurrence. However, we do observe pairs of events whose event measures are highly correlated. The current formulation of the model ignores this source of residual covariance. Hence we describe the events above as “pathology becomes detectable through a *particular* measurement”. Modeling this kind of covariance may increase sensitivity to underlying pathological events and bring \bar{S} closer to the pathology ordering. This necessitates further development both to evaluate dependencies between measurements, which requires larger data sets than we have here, and to accommodate those dependencies within the model. This is a key area for further work, since modeling

measurement dependency further enables relaxation of the one-to-one mapping between measurements and events enabling events informed by multiple measurements and vice versa.

Third, equation 2 requires all sets of measurements X_j to be independent. Here, we treat repeated examinations of the same subject at different time points as independent. This approximation biases \bar{S} towards the progression pattern of individuals with more repeats and underestimates the population variance of the event sequence. This effect may contribute to the greater positional variance in the HD data set compared to fAD. Future work will adapt the estimation procedure to accommodate repeat exams more appropriately, but again this requires larger data sets. Thus, we ignore and simply note the effect for now.

Last, the positional variance diagrams (figure 5) should be interpreted with care, because the diagrams show the posterior distribution of a single event sequence, which is assumed to be common among all patients, rather than the variance of the ordering over the cohort. The positional variance diagram contains contributions from both measurement noise and the underlying population variance. We also note that the posterior distribution contains more information than we explore here. In particular, the positional variance diagrams do not express the covariance in the position of pairs of events, which do exist and may suggest interesting causal relationships. Future work will explore this information.

A key novelty of this work is the formulation of a cohort-level computational model. The event-based model is the simplest of a rich family of related models for future development and study. For instance, the model adapts easily to allow groups of events to occur simultaneously, thus determining natural staging systems directly from the data. Future models might also incorporate a mixture of event sequences to reveal and capture disease sub-types with different progression patterns within a single cohort. Another useful variant of the model might combine multiple measurement types (for instance regional cortical thickness and volume) to inform a single event, which could potentially increase the robustness of event detection. Other variants for future development might retain the temporal information that is currently factored out in order to capture the ordering independently of varying rates of progression across the cohort. This variant could for example identify an appropriate temporal marker against which to measure the timings of each event across the cohort. Wider future applications include characterizations of developmental processes, such as normal aging or the acquisition of language, mathematical ability or other skills.

In conclusion, we have introduced a novel framework for studying and characterizing disease progression and show progression patterns for two important neurodegenerative diseases in unique detail. The simplicity of the model in its current form enables application even on modest data sets, as demonstrated here. More generally, the idea provides rich opportunities for refinement and further development to characterize features of diseases and similar processes.

Acknowledgments

We thank Martin Rossor, Basil Ridha and Natalie Ryan for their contribution to the clinical aspects of the fAD study. We thank Drs Susie Henley and Ed Wild for their contribution to the recruitment and assessment of the Huntington's disease subjects and controls for the London longitudinal study. We also thank all the participants involved in this study. EPSRC (UK) grants EP/D506468/01 and EP/E007748 supported HMF and DCA on this project. MJC was supported by TSB grant M1638A and CBRC grant 168. NCF was funded by the MRC (UK). The data acquisition for this work took place at UCL/UCLH who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research Trust Coordinating centre and has also received equipment funded by the Alzheimer's Research Trust.

References

- Aylward, E.H., Nopoulos, P.C., Ross, C.A., Langbehn, D.R., Pierson, R.K., Mills, J.A., Johnson, H.J., Magnotta, V.A., Juhl, A.R., Paulsen, J.S., 2010. Longitudinal change in regional brain volumes in prodromal Huntington disease. *J Neurol Neurosurg Psychiatry*.
- Bertram, L., Lill, C.M., Tanzi, R.E., 2010. The genetics of Alzheimer disease: back to the future. *Neuron* 68, 270-281.
- Braak, H., Braak, E., 1991. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol* 82, 239-259.
- Carbone, P.P., Kaplan, H.S., Musshoff, K., Smithers, D.W., Tubiana, M., 1971. Report of the Committee on Hodgkin's Disease Staging Classification. *Cancer Res* 31, 1860-1861.
- Cardenas, V.A., Du, A.T., Hardin, D., Ezekiel, F., Weber, P., Jagust, W.J., Chui, H.C., Schuff, N., Weiner, M.W., 2003. Comparison of methods for measuring longitudinal brain change in cognitive impairment and dementia. *Neurobiol Aging* 24, 537-544.

- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39, 1-38.
- Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31, 968-980.
- Dickerson, B.C., Bakkour, A., Salat, D.H., Feczko, E., Pacheco, J., Greve, D.N., Grodstein, F., Wright, C.I., Blacker, D., Rosas, H.D., Sperling, R.A., Atri, A., Growdon, J.H., Hyman, B.T., Morris, J.C., Fischl, B., Buckner, R.L., 2009. The cortical signature of Alzheimer's disease: regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cereb Cortex* 19, 497-510.
- Du, A.T., Schuff, N., Kramer, J.H., Ganzer, S., Zhu, X.P., Jagust, W.J., Miller, B.L., Reed, B.R., Mungas, D., Yaffe, K., Chui, H.C., Weiner, M.W., 2004. Higher atrophy rate of entorhinal cortex than hippocampus in AD. *Neurology* 62, 422-427.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341-355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Segonne, F., Salat, D.H., Busa, E., Seidman, L.J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A.M., 2004. Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14, 11-22.
- Folstein, M.F., Folstein, S.E., McHugh, P.R., 1975. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 12, 189-198.
- Fonteiijn, H.M., Clarkson, M.J., Modat, M., Barnes, J., Lehmann, M., Ourselin, S., Fox, N.C., Alexander, D.C., 2011. An Event-Based Disease Progression model and its application to familial Alzheimer's Disease. *Information Processing in Medical Imaging*.
- Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC.
- Henley, S.M., Wild, E.J., Hobbs, N.Z., Scahill, R.I., Ridgway, G.R., Macmanus, D.G., Barker, R.A., Fox, N.C., Tabrizi, S.J., 2009. Relationship between CAG repeat length and brain volume in premanifest and early Huntington's disease. *J Neurol* 256, 203-212.
- Herrup, K., 2010. Reimagining Alzheimer's disease--an age-based hypothesis. *J Neurosci* 30, 16755-16762.
- The Huntington's Disease Collaborative Research Group, 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72, 971-983.
- Huntington's Study Group, 1996. Unified Huntington's Disease Rating Scale: reliability and consistency. *Mov Disord* 11, 136-142.

- Jack, C.R., Jr., 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.
- Korf, I., 2004. Gene finding in novel genomes. *Bmc Bioinformatics* 5.
- 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.
- Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine* 98, 278-284.
- Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement* 1, 55-66.
- Paulsen, J.S., Langbehn, D.R., Stout, J.C., Aylward, E., Ross, C.A., Nance, M., Guttman, M., Johnson, S., MacDonald, M., Beglinger, L.J., Duff, K., Kayson, E., Biglan, K., Shoulson, I., Oakes, D., Hayden, M., 2008. Detection of Huntington's disease decades before diagnosis: the Predict-HD study. *J Neurol Neurosurg Psychiatry* 79, 874-880.
- Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 56, 303-308.
- Puolamäki, K., Fortelius, M., Mannila, H., 2006. Seriation in paleontological data using Markov Chain Monte Carlo methods. *PLoS Computational Biology* 2, e6.
- Rabiner, L.R., 1989. A Tutorial on Hidden Markov-Models and Selected Applications in Speech Recognition. *Proceedings of the Ieee* 77, 257-286.
- Ridha, B.H., Barnes, J., Bartlett, J.W., Godbolt, A., Pepple, T., Rossor, M.N., Fox, N.C., 2006. Tracking atrophy progression in familial Alzheimer's disease: a serial MRI study. *Lancet Neurol* 5, 828-834.
- Rosas, H.D., Salat, D.H., Lee, S.Y., Zaleta, A.K., Pappu, V., Fischl, B., Greve, D., Hevelone, N., Hersch, S.M., 2008. Cerebral cortex and the clinical expression of Huntington's disease: complexity and heterogeneity. *Brain* 131, 1057-1068.
- Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N., Fox, N.C., 2002. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proc Natl Acad Sci U S A* 99, 4703-4707.
- Shoulson, I., Fahn, S., 1979. Huntington disease: clinical care and evaluation. *Neurology* 29, 1-3.
- Tabrizi, S.J., Langbehn, D.R., Leavitt, B.R., Roos, R.A., Durr, A., Craufurd, D., Kennard, C., Hicks, S.L., Fox, N.C., Scahill, R.I., Borowsky, B., Tobin, A.J., Rosas, H.D., Johnson, H., Reilmann, R., Landwehrmeyer, B., Stout, J.C., 2009. Biological and clinical manifestations of Huntington's disease in the longitudinal TRACK-HD study: cross-sectional analysis of baseline data. *Lancet Neurol* 8, 791-801.
- Tabrizi, S.J., Scahill, R.I., Durr, A., Roos, R.A., Leavitt, B.R., Jones, R., Landwehrmeyer, G.B., Fox, N.C., Johnson, H., Hicks, S.L., Kennard, C., Craufurd, D.,

Frost, C., Langbehn, D.R., Reilmann, R., Stout, J.C., 2011. Biological and clinical changes in premanifest and early stage Huntington's disease in the TRACK-HD study: the 12-month longitudinal analysis. *Lancet Neurol* 10, 31-42.

Thompson, P.M., Hayashi, K.M., de Zubicaray, G., Janke, A.L., Rose, S.E., Semple, J., Herman, D., Hong, M.S., Dittmer, S.S., Doddrell, D.M., Toga, A.W., 2003. Dynamics of gray matter loss in Alzheimer's disease. *J Neurosci* 23, 994-1005.

Thompson, P.M., Mega, M.S., Woods, R.P., Zoumalan, C.I., Lindshield, C.J., Blanton, R.E., Moussai, J., Holmes, C.J., Cummings, J.L., Toga, A.W., 2001. Cortical change in Alzheimer's disease detected with a disease-specific population-based brain atlas. *Cereb Cortex* 11, 1-16.

Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D., Richardson, E.P., Jr., 1985. Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol* 44, 559-577.

Xu, Y., Jack, C.R., Jr., O'Brien, P.C., Kokmen, E., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, R.G., Petersen, R.C., 2000. Usefulness of MRI measures of entorhinal cortex versus hippocampus in AD. *Neurology* 54, 1760-1767.

Figure captions

Figure 1

Conceptual overview of the event-based disease progression model. **(a)** Idealized time courses of event measures during the disease. The data X_j from patient j provides a snapshot of disease progression (solid black lines). These time courses are simply examples that match the assumptions of the model and play no explicit role in fitting the model. **(b)** The left panel shows simulated measurements for controls (green) and patients (red). Patients include cases in which the event has and has not occurred; controls only “not occurred”. The right panel shows the histograms of the measurement for all patients (red) and all controls (green). We fit a mixture of a Gaussian and a uniform distribution (dotted lines in right panel of **(b)**) to all subjects' measurements of one event to estimate $p(x_{ij} | E_i)$ (the uniform component) and $p(x_{ij} | \neg E_i)$ (the Gaussian component) for each event in each patient. These likelihoods are visualized in **(c)** for the four event measures and six

representative patients. The events are ordered according to the underlying event sequence, whereas the patients are ordered according to their disease stage to facilitate visualization. Note that $p(x_{ij} | E_i)$ (right panel in **c**) is the same for all patients within each event, except for a cut-off, before which $p(x_{ij} | E_i)$ is zero. This reflects the fact that these components are modeled by uniform distributions, whose upper bounds are set to the mean of the respective $p(x_{ij} | \neg E_i)$ (see section 2.3). **(d)** The data likelihood supports a MCMC algorithm, which samples the posterior distribution on the ordering S to provide a list of T candidate orderings with the most likely appearing the most often (see left panel **d**). Finally, we compute the characteristic ordering \bar{S} , which orders the events by their average position in the MCMC samples, as in figures 2 and 3. The positional variance diagram shows the histogram of positions of each event over the MCMC samples, which gives insight into the uncertainty on \bar{S} .

Figure 2

Plots of the log-likelihood during the greedy ascent phase (panel **(a)**) and the MCMC phase (panel **(b)**) of the algorithm. All 10 iterations of the greedy ascent phase are initialized randomly, yet converge on a common solution within 2000 iterations. The MCMC trace shows good mixing, while remaining relatively close to the maximum-likelihood solution.

Figure 3

Event-based disease-progression model for the fAD cohort. The upper panel shows the characteristic ordering \bar{S} . The horizontal axis is position in \bar{S} ; vertical spacing is arbitrary and simply separates the positions of the icons for visualization purposes. The lower panel shows snapshots of the reconstructed time course of the spread of atrophy through the brain in fAD. These snapshots show in red the cumulative set of regions in which significant atrophy has occurred up to this stage. The events that have occurred since the last snapshot are listed beneath each snapshot in order of appearance. The names of the regions are consistent with the FreeSurfer nomenclature (Desikan et al., 2006). The red events in the upper panel are the clinical events. In particular, B and C represent transition to clinical MCI and AD respectively. We include event A (pre-MCI) largely as a sanity check, as it must occur before all other events.

Figure 4

Event-based disease-progression model for the HD cohort. See figure 3 for further information.

Figure 5

Positional variance of the event sequence for both diseases. Each row shows the logarithm of the frequency with which a region occupies each position in the event sequence over the set of samples from the MCMC. The regions are ordered according to

\bar{S} . The uniform square at the bottom right of the fAD diagram contains events that show no substantial deviation from controls.

Figure 6

Staging results for all follow-up scans in all patients in the fAD cohort. We use the event-based disease progression model to compute the most likely disease stage for each scan. Each separate trajectory joins the points representing each individual scan from one individual patient. The horizontal axis indicates the number of the follow-up scan. Within patients, later follow-up scans generally correspond to later stages in the model.

Figure 7

Comparison of orderings of left and right events in a model containing events for regions in both the left and the right hemisphere of the fAD cohort. We fit separate models for both hemispheres. Each point plots the position of a left-hemisphere event, relative to all left-hemisphere events, against the position of the corresponding right-hemisphere event relative to all right-hemisphere events. The horizontal and vertical error bars indicate the positional standard deviation over all MCMC samples for the left and right hemisphere respectively.

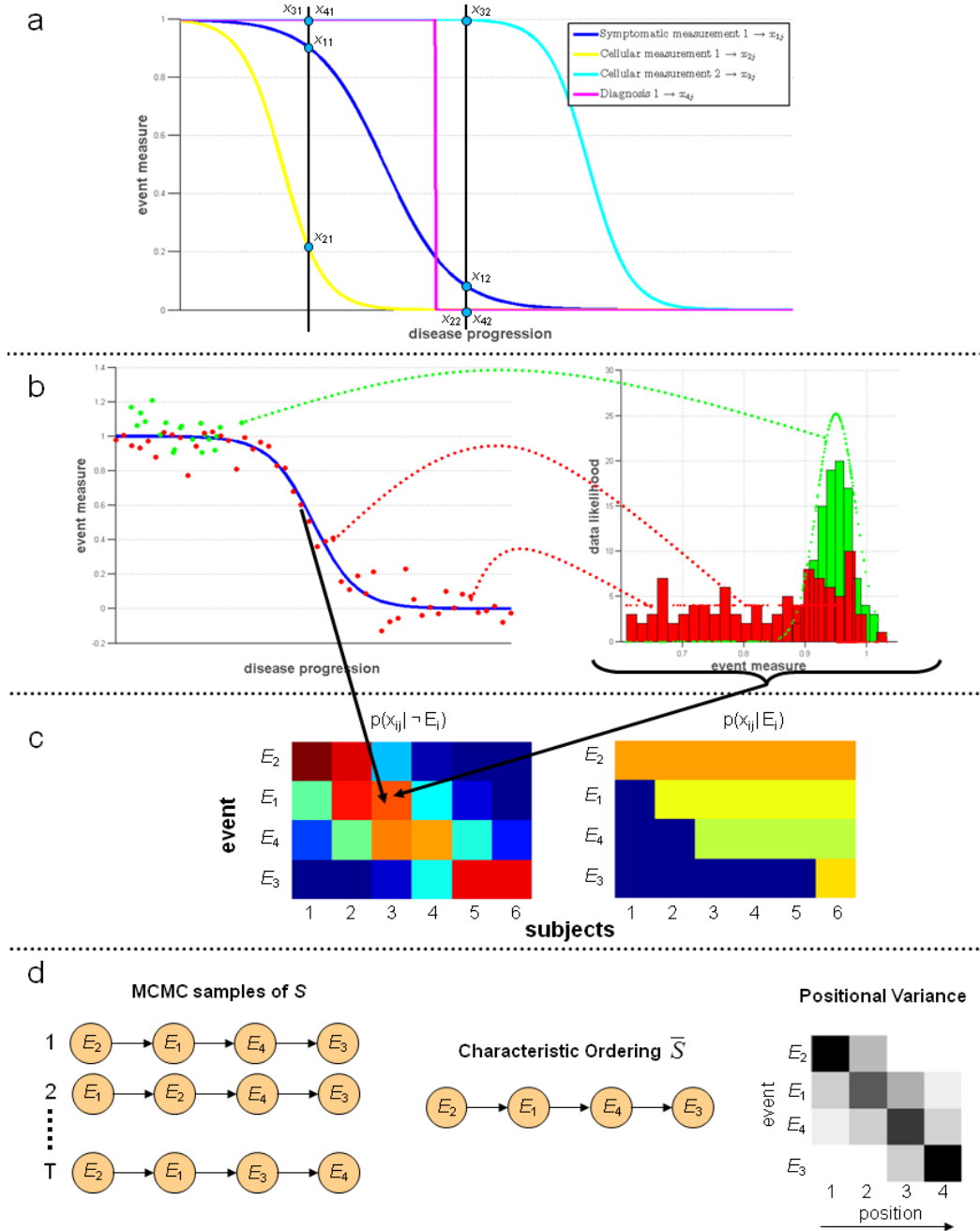
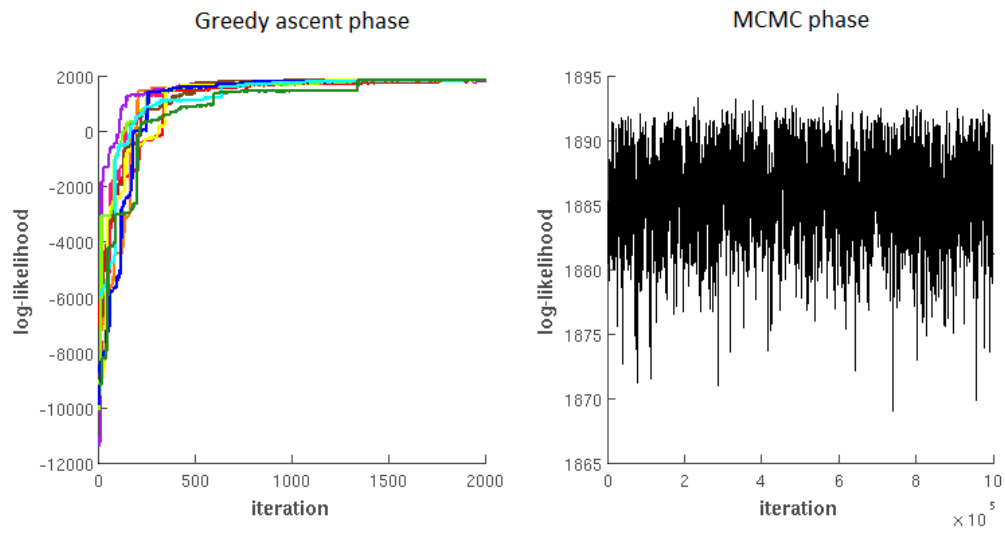


Fig. 1

**Fig. 2**

ACCEPTED

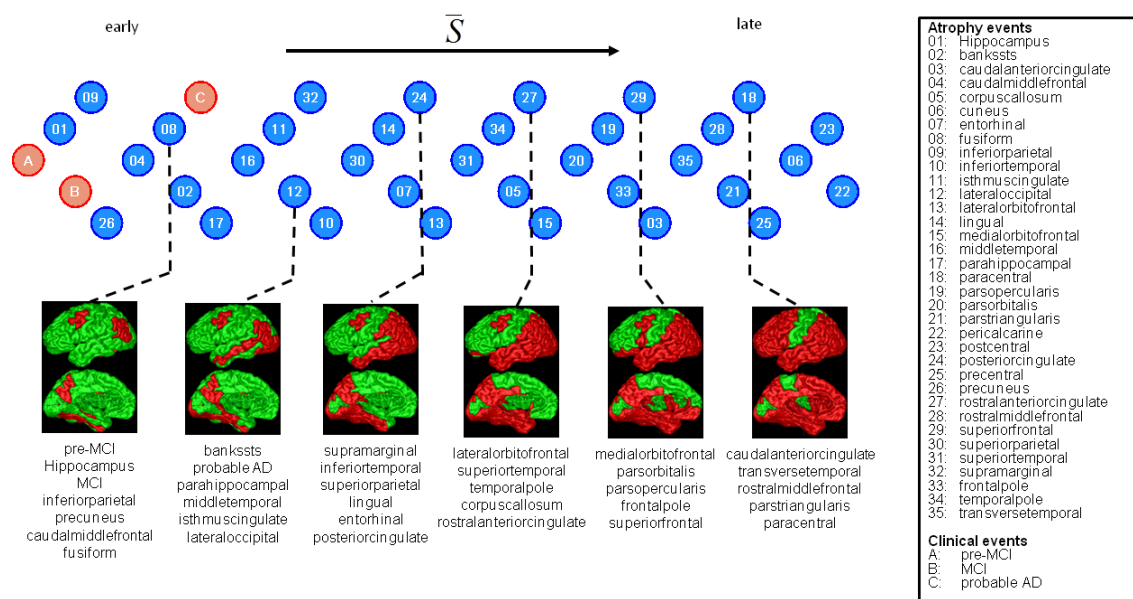


Fig. 3

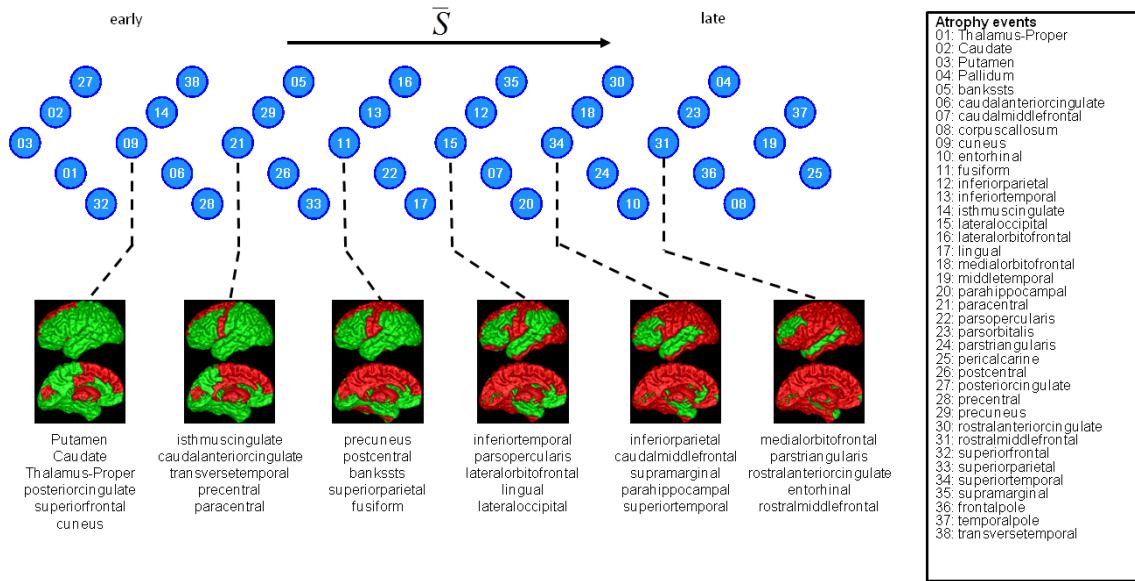


Fig. 4

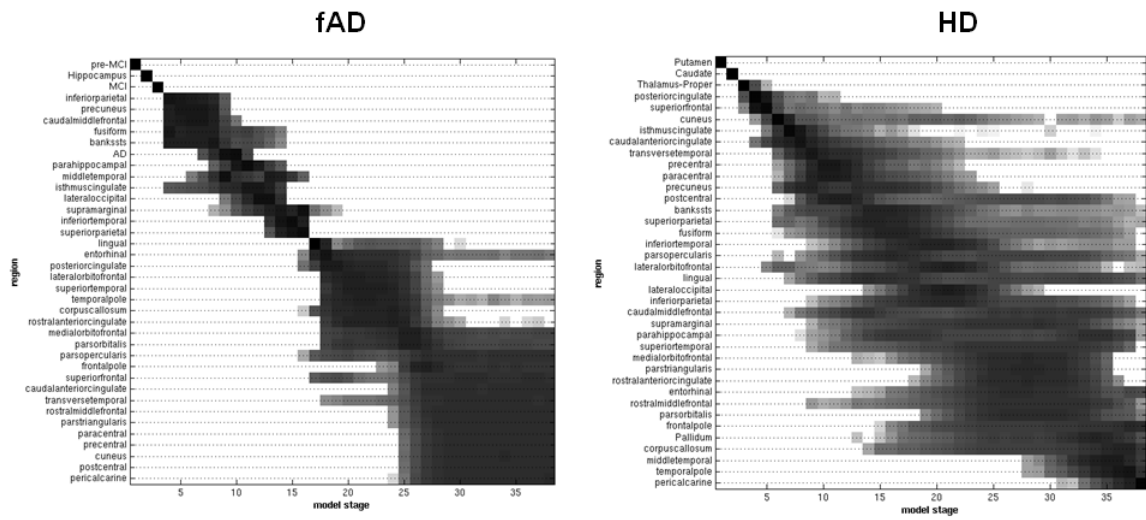


Fig. 5

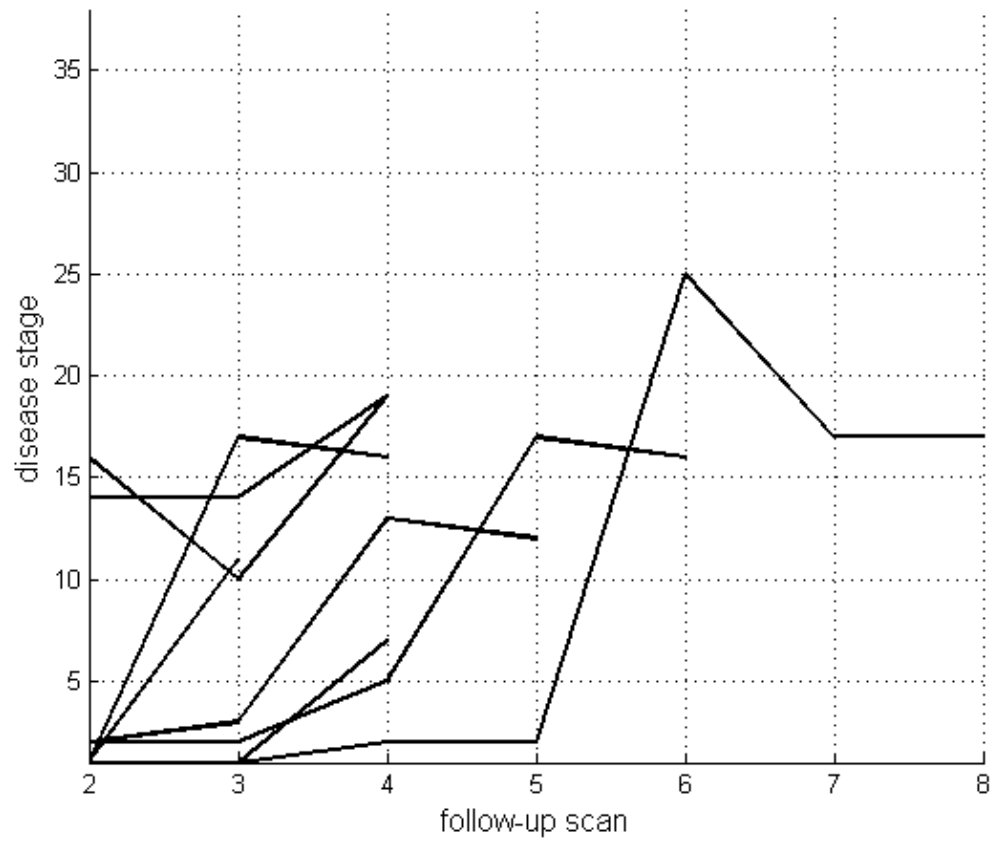


Fig. 6

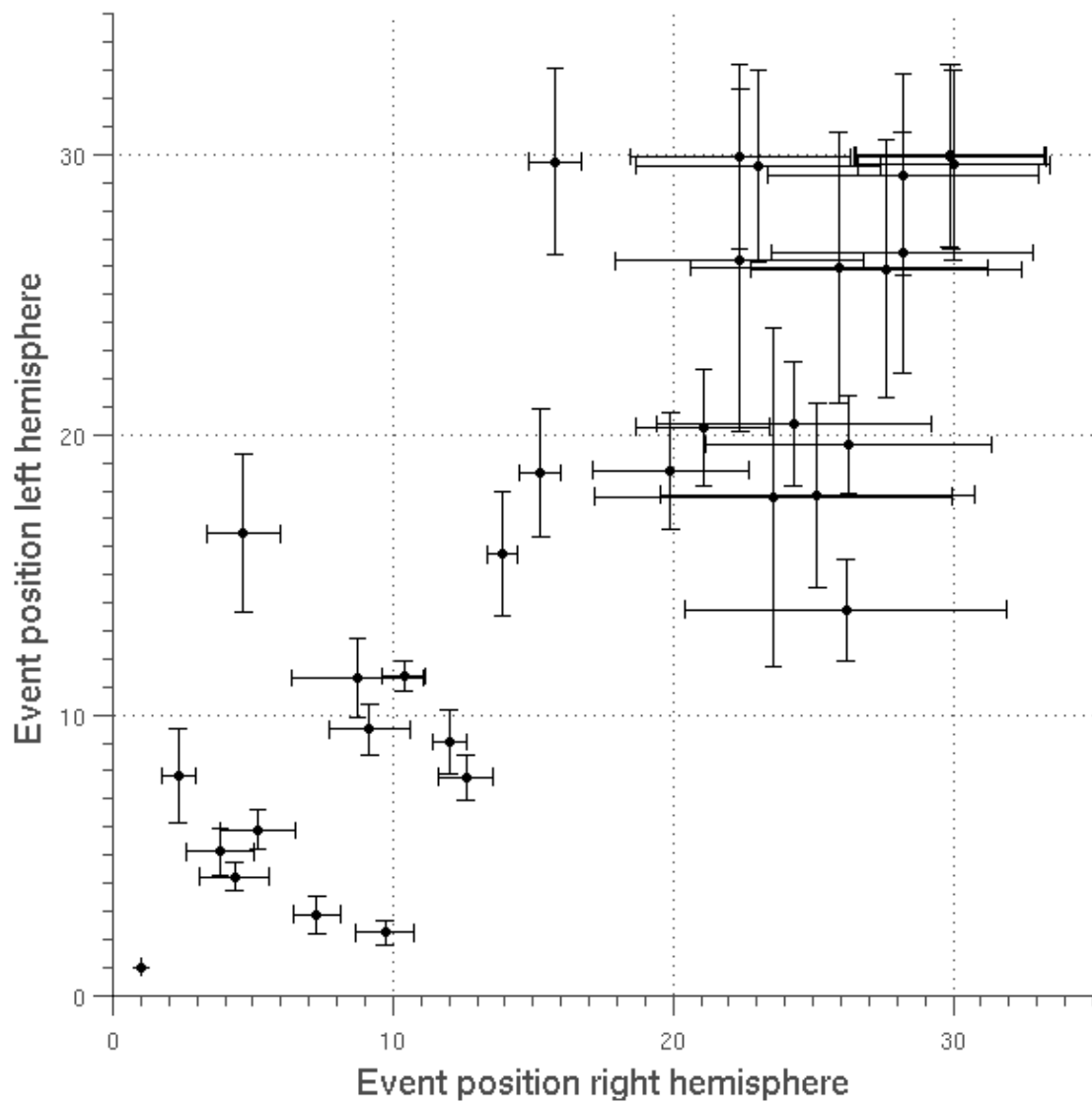


Fig. 7