

## FEATURED ARTICLE

# Sequences of cognitive decline in typical Alzheimer's disease and posterior cortical atrophy estimated using a novel event-based model of disease progression

Nicholas C. Firth<sup>1,2</sup> | Silvia Primativo<sup>3</sup> | Emilie Brotherhood<sup>2</sup> | Alexandra L. Young<sup>1,4</sup> | Keir X.X. Yong<sup>2</sup> | Sebastian J. Crutch<sup>2</sup> | Daniel C. Alexander<sup>1,5</sup> | Neil P. Oxtoby<sup>1</sup>

<sup>1</sup>Centre for Medical Image Computing, Department of Computer Science, UCL, London, WC1E 6BT, UK

<sup>2</sup>Dementia Research Centre, UCL Queen Square Institute of Neurology, UCL, London, WC1N 3BG, UK

<sup>3</sup>Department of Human Science, LUMSA University, Rome, Italy

<sup>4</sup>Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College, London, UK

<sup>5</sup>Clinical Imaging Research Centre, National University of Singapore, Singapore

## Correspondence

Neil P. Oxtoby, Centre for Medical Image Computing, Department of Computer Science, UCL, London, WC1E 6BT, UK.  
Email: n.oxtoby@ucl.ac.uk

## Funding information

UCL; EPSRC, Grant/Award Number: EP/M006093/1; Alzheimer's Research UK; Weston Brain Institute; NIHR; UKRI, Grant/Award Number: MRC MR/S03546X/1

[The copyright line for this article was changed on June 12, 2020 after original online publication.]

## Abstract

**Introduction:** This work aims to characterize the sequence in which cognitive deficits appear in two dementia syndromes.

**Methods:** Event-based modeling estimated fine-grained sequences of cognitive decline in clinically-diagnosed posterior cortical atrophy (PCA) ( $n = 94$ ) and typical Alzheimer's disease (tAD) ( $n = 61$ ) at the UCL Dementia Research Centre. Our neuropsychological battery assessed memory, vision, arithmetic, and general cognition. We adapted the event-based model to handle highly non-Gaussian data such as cognitive test scores where ceiling/floor effects are common.

**Results:** Experiments revealed differences and similarities in the fine-grained ordering of cognitive decline in PCA (vision first) and tAD (memory first). Simulation experiments reveal that our new model equals or exceeds performance of the classic event-based model, especially for highly non-Gaussian data.

**Discussion:** Our model recovered realistic, phenotypical progression signatures that may be applied in dementia clinical trials for enrichment, and as a data-driven composite cognitive end-point.

## KEYWORDS

Alzheimer's disease, ceiling, cognitive decline, dementia, disease progression model, effect, floor, kernel density estimate, non-Gaussian, nonparametric mixture model, posterior cortical atrophy

## 1 | INTRODUCTION

Alzheimer's disease (AD) is the leading contributor to the global dementia epidemic. With no disease-modifying therapeutic currently available and over 99% of clinical trials concluding without evidence of efficacy for their putative therapy,<sup>1–5</sup> there is an immediate urgency to improve our understanding of AD, and dementia in general, to inform these efforts.

It is estimated that measurable changes in biomarkers can occur decades before a dementia diagnosis is made, such as amyloid abnor-

mality in typical (memory-led) AD.<sup>6</sup> This makes dementia particularly challenging to study because it is near-impossible to discern precisely when a given change occurs from a cohort of diagnosed or prodromal individuals, without following them for a prohibitively long period of time.

The event-based model<sup>7,8</sup> is a computational algorithm specifically designed to meet this challenge, uniquely using only a cross-sectional sample such as the baseline visit in a study. The event-based model estimates the ordered sequence of abnormality in a set of biomarkers by combining severity information across biomarkers and individuals,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Alzheimer's & Dementia* published by Wiley Periodicals LLC on behalf of Alzheimer's Association.

without reference to a given individual's clinical status. It has been applied in both sporadic and familial AD,<sup>8–11</sup> Huntington's disease,<sup>12,13</sup> and progressive multiple sclerosis.<sup>14</sup> Recently, the event-based model idea has been extended to a new algorithm for finding data-driven subtypes of disease,<sup>15</sup> demonstrated in AD and frontotemporal dementia. These investigations, and others,<sup>16</sup> focused largely on neuroimaging biomarkers, made possible by the increasing availability of large datasets such as from the Alzheimer's Disease Neuroimaging Initiative and the Dominantly Inherited Alzheimer Network.

In contrast to neuroimaging, most analyses of complex cognitive datasets have relied on traditional statistical approaches rather than data-driven methods. Methods of detecting cognitive change are important both for improving disease characterization and prognosis in affected individuals, and for detecting and predicting change in individuals who are asymptomatic and at-risk (of sporadic disease) or presymptomatic (have a familial/genetic disease).<sup>17</sup> Optimizing methods for analyzing cognitive change is especially important in the context of clinical trials, because cognitive and functional outcomes are currently the only accepted means for proving efficacy of a drug. This is relevant to both symptomatic trials and secondary prevention trials, because neuropsychological tests may be sensitive to dementia between 10 and 17 years before diagnosis.<sup>18</sup>

Evaluation of longitudinal change within and across cognitive domains presents a number of specific challenges. First, performance across cognitive tasks is not independent. General factors (eg, disease severity) and collateral deficits (eg, visuo-perceptual problems limiting performance on a visual memory test) can influence testing across domains. Second, cognitive profiles across tasks are often described qualitatively because test properties and normative samples differ across tasks. Third, the psychometric shape of tests can differ markedly. Tests involving graded difficulty yield relatively linear score distributions among healthy control participants, whereas other tests may yield skewed, highly non-Gaussian score distributions owing to an excess of very-easy or very-difficult items. These properties influence the likelihood of clinical populations showing ceiling or floor effects at any given point in their disease progression. Fourth, practice effects mask longitudinal change, for example, test familiarity and/or reduced anxiety may conceal evidence of cognitive instability or decline.<sup>19</sup>

We are motivated to understand disease progression in Alzheimer's and dementia, with a focus on impact for interventional trials and in the clinic. International Working Group criteria now include explicit definition of atypical forms of AD,<sup>20,21</sup> of which posterior cortical atrophy (PCA) is acknowledged to be one of the most common.<sup>22,23</sup> It is of fundamental importance to understand disease progression in both typical and atypical AD if the field is to advance. PCA is a clinico-radiological syndrome characterized by progressive decline in visual processing and other posterior cognitive functions, relatively intact memory and language in the early stages, and atrophy of posterior brain regions.<sup>20,21</sup> PCA is most commonly caused by AD, with greater presence of amyloid plaque and/or neurofibrillary tangles in the posterior cortices than in individuals having the more-typical, amnesic presentation.<sup>24,25</sup> Detailed longitudinal studies of cognitive change in PCA are understandably rare.<sup>26</sup>

## RESEARCH IN CONTEXT

1. Systematic review: We reviewed the literature using PubMed and found no data-driven estimation of cognitive decline in posterior cortical atrophy.
2. Interpretation: Our data-driven results reveal new understanding into the sequence of detectable cognitive decline in posterior cortical atrophy and tAD, including the first direct comparison. Our novel method produces a fine-grained prognostic tool useful for staging individuals based on their cognitive profile, and for predicting subsequent decline.
3. Future directions: We foresee multiple avenues for future work. (1) Validation of our model using larger datasets in tAD. (2) Application of these models to clinical trials in tAD and posterior cortical atrophy (enrichment; and data-driven composite cognitive end-points). (3) Wider application of our new method in other neurodegenerative diseases and beyond neuropsychological test scores, including data-driven subtyping of diseases. (4) Working toward making KDE mixture modeling a more general statistical methodology.

**TABLE 1** Demographics of participants

	PCA (n = 94)	tAD (n = 61)	HC (n = 23)
Age (years)	64.2 (7.9)	65.8 (7.7)	60.2 (5.7)
Gender (m:f)	35:59	38:23	11:12
MMSE, mean (std)	21.4 (5.1)	19.6 (4.9)	29.5 (0.7)
N visits, mean (std)	2.5 (1.3)	1.6 (0.6)	1.8 (0.4)

Abbreviations: HC, healthy control; MMSE, Mini-Mental State Examination; PCA, posterior cortical atrophy; tAD, typical Alzheimer's disease.

Here we aimed to estimate the sequence of cognitive decline in typical AD (tAD) and PCA. This necessitated development of a new event-based model designed specifically to handle highly non-Gaussian input data for data-driven assessment of cognitive decline. We validated our new model in simulation experiments (documented in Supplementary Material) before applying to patient data from our cohorts.

## 2 | MATERIALS AND METHODS

### 2.1 | Participants in the two dementia cohorts

Patients with a clinical diagnosis of PCA (n = 94) or tAD (n = 61) were recruited between October 2005 and June 2016 at the UCL Dementia Research Centre in London. Healthy controls (HC) (n = 23) were sampled from the Young Onset Alzheimer's Disease study (Table 1) at the same center. Our control group is younger than both patient groups (Mann-Whitney U test,  $P < .05$ ). Patients were recruited via attendance at the Cognitive Disorder Clinic at the National Hospital

for Neurology and Neurosurgery, or were recruited by individual referral from neurologists to whom they expressed interest in taking part in observational research. All PCA patients met both Tang-Wai et al.<sup>24</sup> and Mendez et al.<sup>27</sup> criteria based on available information at baseline and expert retrospective clinical review by consultant neurologists and neuropsychologists with expertise in cognitive neurology. PCA participants were excluded if they also met criteria for another neurodegenerative syndrome, thus our PCA group fulfils consensus criteria for PCA-pure.<sup>23</sup> Patients with PCA and patients with tAD fulfilled research criteria for probable AD.<sup>28,29</sup> Examples of excluded syndromes include dementia with Lewy bodies, corticobasal degeneration, and prion disease; with clinical features such as visual hallucinations, pyramidal signs, reduplicative phenomena, parkinsonism, dystonia, myoclonus, and ataxia. Additionally, tAD patients did not fulfil clinical criteria for logopenic variant of primary progressive aphasia<sup>30</sup> nor frontal variant AD.<sup>20</sup> Event-based models (see Section 2.3.1) were fit to baseline data for PCA and tAD separately. Follow-up visits were used to assess longitudinal self-consistency of each model: the mean  $\pm$  std number of visits per patient is  $3.1 \pm 1.1$  (range 2–6) for PCA and  $2.2 \pm 0.4$  (range 2–3) for tAD. We investigated controlling for visual acuity to avoid confounding visual cognition, but we found it impossible to separate acuity from cognition in our PCA patients – even on acuity tests intended to be robust to deficits in visual processing. Information on patient medication was not routinely collected in our cohorts.

Data collection was approved by the National Research Ethics Service Committee London (UK National Health Service Health Research Authority). All participants provided written informed consent according to guidelines established by the Declaration of Helsinki.

## 2.2 | Neuropsychological tests

We employed a battery of tests that are routinely used in the clinic<sup>31–35</sup> and in pharmacological<sup>36</sup> and non-pharmacological<sup>37</sup> clinical trials involving PCA. We used the same battery of neuropsychological tests on each cohort, which allows direct comparison of cognitive decline across the two dementia syndromes. The battery includes assessments of episodic and working memory, visuo-perceptual and visuo-spatial processing, arithmetic, and general cognition. The full list of tests and the primary cognitive domain tested by each is shown in Table 2, along with abbreviations used in the results section. Descriptive statistics of scores for each test and per patient group are provided in Table S1. We found no significant age-related effects in any of the cognitive tests, which reassures us that group differences are due to disease.

## 2.3 | Statistical analysis

### 2.3.1 | Event-based model

The event-based model<sup>7</sup> is designed to estimate a data-driven, probabilistic sequence of biomarker “events” that represents an underlying cumulative process, using a cross-sectional set of observations. These can be any biomarkers. In the context of neurodegenerative

**TABLE 2** Cognitive domains tested by our neuropsychological test battery

Cognitive domain	Neuropsychological test	Abbreviation
Vision	A Cancellation time (time)	A Cancel (time)
	A Cancellation time (number missed)	A Cancel (n Miss)
	VOSP Fragmented Letters	Fragmented Letters
	VOSP Dot Counting	Dot Count (n correct)
	Efron Shape Discrimination	Shape Discrimination
	VOSP Object Decision	Object Decision
Memory	Short Recognition Memory Test (Words)	SRMT (W)
	Short Recognition Memory Test (Faces)	SRMT (F)
	Paired Associate Learning	PAL
	Digit Span (Forwards, total)	Digit Span (F)
	Digit Span (Forwards, maximum)	Digit Span (F Max.)
	Digit Span (Backwards, total)	Digit Span (B)
	Digit Span (Backwards, maximum)	Digit Span (B Max.)
General cognition	Mini Mental State Examination	MMSE
Arithmetic	Graded Difficulty Arithmetic (addition)	GDA (add)
	Graded Difficulty Arithmetic (subtraction)	GDA (sub)
	Graded Difficulty Arithmetic (total)	GDA (tot)

<sup>a</sup>For further details on many of these tests, we refer the reader to [31–37]

diseases, an event corresponds to a group-level statistical deviation from normality/health (defined by data from controls) toward abnormality/disease (defined by data from patients), with the full sequence of events representing the cumulative effects of neurodegenerative disease progression. The ordering of events is determined probabilistically, in a data-driven manner, by pooling biomarker severity (event probability) across individuals. Conceptually, higher prevalence corresponds to an earlier position in the sequence. The event-based model estimates both the sequence and uncertainty in the sequence. Event probability is a function of the likelihood of an event having occurred  $Pr(x_{ij}|E_i)$  (“post-event”), or having not occurred  $Pr(x_{ij}|\neg E_i)$  (“pre-event”), via a two-component univariate mixture model (see Section 2.3.2). Assuming independent observations and biomarkers, the likelihood of an ordered sequence  $S$  is [8]

$$Pr(X|S) = \prod_{j=1}^N \left[ \sum_{m=0}^M \left\{ \prod_{i=1}^m Pr(x_{ij}|E_i) \prod_{i=m+1}^M Pr(x_{ij}|\neg E_i) \right\} \right] \quad (1)$$

where measurements  $x \in X$  come from  $i \in M$  event markers and  $j \in N$  individual samples, such as participants in a disease cohort.

In the absence of prior information, a uniform prior distribution over sequences is used and the characteristic ordering  $\hat{S}$  is the sequence that maximizes the likelihood  $Pr(X|S)$  in (1). This requires a search among the  $M!$  possible sequences, which quickly explodes to preclude an exhaustive search when  $M$  exceeds  $\sim 10$ , necessitating approximation of this maximization. The search is performed using a combination of multiply initialized gradient ascent and Markov chain Monte Carlo sampling.

### 2.3.1.1 | Patient staging

An individual sample  $X_j$  (vector of measurements across biomarkers  $i$  for an individual  $j$ ) is staged within a given sequence via the data likelihood. For example, by finding the stage  $m$  that maximizes the individual likelihood<sup>9</sup>:

$$\arg \max_m Pr(X_j|S, m) = \arg \max_m \prod_{i=1}^m Pr(x_{ij}|E_i) \prod_{i=m+1}^M Pr(x_{ij}|\neg E_i). \quad (2)$$

## 2.3.2 | Mixture modeling and our new event-based model

Disease severity varies among patients, so we fit a two-component mixture model to determine event probability for each biomarker (cognitive test scores in this work). This probabilistic assignment of individuals into two subgroups – pre-event and post-event – allows patients to have combinations of normal and abnormal observations across biomarkers, which in turn allows for the sequence of events to be estimated. Previous event-based models have incorporated parametric mixture models – most commonly a Gaussian mixture model, which works well for many data types including imaging markers that have been the primary application of the event-based model to date. However, a Gaussian mixture model can be highly inappropriate for skewed data, such as the cognitive test scores considered here. Indeed, we show in Supplementary Material that Gaussian mixture modeling produces erroneous event sequences, necessitating our alternative, nonparametric mixture model proposed here – see Figures S3, S4, and S5.

We emphasize that events are inherently probabilistic. While we use the language of events having “occurred” or not, this is in an explicitly probabilistic sense. This is one of the event-based model's key benefits – that explicit biomarker cutpoints are not required. We now discuss how the mixture modeling is used to calculate event probabilities.

### 2.3.2.1 | Parametric, Gaussian mixture modeling

Two-component Gaussian mixture model fitting involves estimating the mean and variance of each component, and a mixture weight. We initialized the pre-/post-event components using the diagnostic labels control/diseased. We initialized equal mixture weights of 0.5 and would terminate the fit if either weight exceeded the range [0.1,0.9] to ensure that components do not vanish when the data from diagnostic groups overlap considerably. Parameters were optimized to minimize the negative log-likelihood of the data given the mixture model, using

the sequential least-squares programming (SLSQP) algorithm. These choices of constraints, initial parameters, and the SLSQP algorithm are similar to the classic event-based model.<sup>8,9</sup>

### 2.3.2.2 | Non-parametric, kernel density estimation mixture modeling

Kernel density estimation (KDE) is a non-parametric method for estimating a probability density. The KDE estimate  $\hat{f}(x)$  of a function  $f(x)$  with an independent and identically distributed sample  $\{x_j\}$  drawn from a distribution with an unknown density is given by

$$\hat{f}(x) = \frac{1}{Nh} \sum_{j=1}^N K\left(\frac{x - x_j}{h}\right), \quad (3)$$

where  $K$  is a non-negative zero-mean kernel function that integrates to unity and  $h$  is a positive smoothing factor called a bandwidth. With an appropriate choice of  $K$ , KDE naturally extends to multivariate density estimation. In this work we estimate mixture model components using the KDE implementation in `scikit-learn`,<sup>38</sup> with a default (Gaussian) kernel and default parameters for all values except the bandwidth (see below). The choice of kernel function  $K$  is not dependent upon the datatype of the biomarker being modeled. For example, there is no restriction against using a smooth kernel function when modeling discrete-valued cognitive test scores.

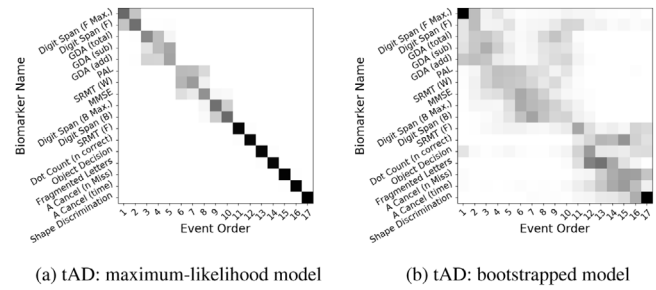
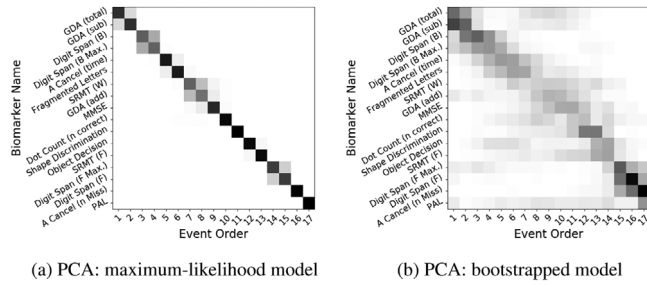
The KDE approach enables data-driven estimation of probability densities that do not follow a parametric distribution, which is particularly useful for cognitive test scores that are susceptible to ceiling effects and floor effects. Our novel mixture modeling algorithm incorporates non-parametric mixture modeling using KDE components. As for our Gaussian mixture modeling, we initialized the mixture model components using the labeled data (HC, diseased), and start with equal mixture weights that are restricted to values within the range [0.1,0.9] to ensure that components do not vanish. We choose to fix the biomarker bandwidth  $h$  using Scott's normal reference rule<sup>39</sup> applied jointly on both groups. A variable bandwidth would be sensible for KDE of highly skewed data with sparsely populated tails. Our full algorithm is given in the Supplementary Material and code is available at [https://github.com/noxtooby/kde\\_ebm\\_open](https://github.com/noxtooby/kde_ebm_open).

## 2.3.3 | Cross-validation

We performed cross-validation of our event-based models by re-estimating each full model (event distributions and maximum-likelihood sequence) on 100 bootstrap samples (sampling with replacement). The resulting bootstrapped model tends to overestimate positional variance in the event sequence.

## 3 | RESULTS

Here we present results from our experiments on real data in two dementia syndromes using our new event-based model incorporating



**FIGURE 1** Data-driven sequence of cognitive decline in posterior cortical atrophy (PCA): (a) Maximum-likelihood model; (b) Bootstrapped model. The bootstrapped model overestimates positional variance in the sequence (see Materials and Methods, Section 2). Grayscale intensity represents the proportion (0 in white, 1 in black) of the posterior Markov chain Monte Carlo samples in which events (y-axis) appear in a particular position (x-axis) in the sequence

**FIGURE 2** Data-driven sequence of cognitive decline in typical Alzheimer's disease (tAD): (A) Maximum-likelihood model; (B) Bootstrapped model. See Figure 1 for further explanation of positional variance diagrams

a KDE-component mixture model. In Figures S6–S8 we present a comparison with the existing state of the art — the classic event-based model<sup>9</sup> incorporating a Gaussian mixture model. Detailed performance evaluation results from our simulation experiments are in Figures S3–S5, where we show that our new method outperforms the state of the art method in situations where the underlying biomarker distributions are similar to those observed from cognitive tests.

### 3.1 | Sequence of cognitive decline in PCA

Figure 1 is a visualization of the probabilistic sequence of detectable cognitive decline due to PCA that was estimated by our data-driven method, using the set of neuropsychological tests in our battery. The posterior positional variance shows the model's confidence (left-to-right) in the ordering (top-to-bottom): narrow, dark sections of low positional variance show high confidence in the ordering. Figure 1A (left) shows the maximum-likelihood model and Figure 1B (right) shows a more conservative estimate of the variability in the sequence using bootstrapping.

In agreement with clinical knowledge, our results suggest that measures of visual processing become abnormal earlier in PCA patients (A Cancellation (time) and VOSP Fragmented Letters), while the model estimates relatively late deficits in episodic memory (Short Recognition Memory Test [SRMT] and Paried-Associate Learning [PAL]), although in the case of PAL this may be influenced by missing data (see Supplementary Material). The location of PAL does not affect the ordering of other events. Deteriorating performance on an arithmetic test (GDA) was estimated to be the earliest event, although the bimodal nature of the positional density under bootstrapping (Figure 1B, top rows) may reflect heterogeneity within PCA. Tasks having higher demands on working memory (Digit Span backward) appeared earlier than simpler short-term memory tasks (Digit Span forward), although the latter is less certain as indicated by the weakly bimodal positional density under bootstrapping.

Our model produces results more consistent with intuition and expectations than the classic event-based model (Figures S3–S9). Specifically, in our model the highly correlated subscores (total items correct and maximum span) of both the forward (F, F max.) and backward (B, B max.) Digit Span tests appear in successive positions in the event sequence, as expected.<sup>40</sup> The same cannot be said of the classic event-based model incorporating a Gaussian mixture model (Figure S6) where these subscores are almost maximally separated in the sequence. In the case of GDA subscores, the situation may appear to be reversed (our model separates correlated subscores, while the classic event-based model does not) until one considers the aforementioned bimodal nature of the bootstrapped density for GDA in Figure 1B. This warrants further investigation.

### 3.2 | Sequence of cognitive decline in tAD

Figure 2 is a visualization of our method's data-driven sequence of detectable cognitive decline due to tAD, using the same set of neuropsychological tests as for PCA in Figure 1.

In agreement with clinical knowledge, our results suggest that tAD patients show early deterioration in episodic memory (PAL and SRMT) with visual processing deficits clearly confined to the end of the sequence, even in the more-variable bootstrapped model of Figure 2B. This is largely a reversal of the sequence estimated for PCA, with the exception of GDA which was among the earliest events in both dementia syndromes (see Discussion in Section 4). Digit Span was also involved very early, although the bootstrapped model contains bimodal positional density either side of episodic memory (PAL and SRMT), which may reflect heterogeneity within tAD.

### 3.3 | Patient staging and longitudinal self-consistency of models

We use patient staging to assess model self-consistency. A valid model consistent with the progressive nature of dementia will produce non-decreasing disease stages for individuals from baseline to follow-up(s), within model uncertainty. Our disease progression

models are trained on baseline data, with longitudinal self-consistency assessed on follow-up data (a separate test set, albeit from the same individuals). Described mathematically in Section 2, within each dementia syndrome we assigned an event-based model stage to each individual using Equation 2, which is akin to aligning an individual's cognitive profile (test scores) with the maximum-likelihood model. We emphasize that these are self-consistency checks within a model which are not useful for direct model comparison purposes.

Figure 3 summarizes the distribution of event-based model stages in PCA (left) and tAD (right) as histograms. As expected, HC generally perform better across all tests and are assigned to early stages. Commensurately, patients generally exhibit poorer performance and are assigned later stages — mostly very late stages. The only notable exception is a small fraction of tAD patients who have been assigned a very early event-based model stage in Figure 3B. These four patients have mild symptoms (MMSE score  $\geq 24$ , mean 25.7) and all performed considerably better than other patients on Digit Span Forwards and GDA (the earliest events), which dominated the mild abnormality in later events including MMSE.

Quantitatively, our models show high longitudinal self-consistency. We calculate the proportion of pairs of visits within individuals having non-decreasing model stage, within model uncertainty (positional variance). In PCA we found 244 of 262 available follow-up combinations to be consistent (93%). In tAD we found 34 out of 39 (89%). These self-consistency results are not influenced by baseline stage (early vs late) because the number of visits per patient is comparable across baseline stages.

## 4 | DISCUSSION

We have revealed fine-grained representations of deterioration across cognitive domains in two dementia syndromes: PCA and tAD. This was enabled by a novel event-based model for estimating data-driven sequences of cognitive decline in neurodegenerative diseases, as well as uncertainty in the sequences.

We estimated sequences of cognitive decline that are broadly consistent with clinical criteria for, and reports of progression in, PCA and tAD. This supports the use of our method in biological applications where high-precision staging is required, such as for screening in clinical trials. Decline in the five-principle visual tests was estimated to occur earlier (A cancellation time, Fragmented Letters, Dot Counting, Shape Discrimination), or at a comparable stage (Object Decision), in PCA relative to tAD (mean  $5.6 \pm 4.2$  positions earlier, among 17 possible positions). Likewise, decline in the three-principle episodic memory tasks (short Recognition Memory Test for Words and Faces, and the Paired Associate Learning test) was estimated to occur somewhat earlier in tAD than PCA (mean  $4.3 \pm 5.9$  positions earlier). Arithmetic performance was estimated to decline early in both patient groups, which might relate to this graded task being designed both to obtain a normal (Gaussian) distribution of scores and also to avoid floor and ceiling effects. Furthermore, the timed nature of the Graded Difficulty Arithmetic (GDA) test makes it sensitive to general decline in attention and

processing speed (affected in both dementia syndromes), in addition to acalculia.

We validated our models by assessing longitudinal self-consistency of patient staging on a separate test set: the follow-up data in each cohort. In PCA, we observed 93% longitudinal consistency (non-decreasing patient stage at follow-up), and in tAD we observed 89%.

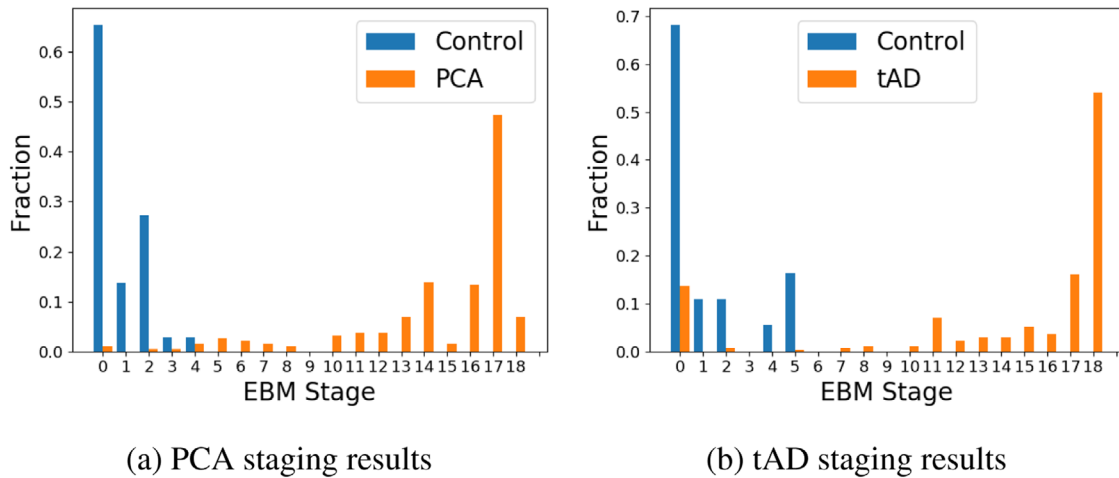
We further assessed the robustness of our results by comparing the maximum-likelihood models with their bootstrapped counterparts, the latter of which tends to overestimate uncertainty. The estimated sequences are robust, with the 2D positional density maps remaining consistently close to the diagonal.

A number of previous studies in PCA found broadly supportive results. Our finding that visual and visuospatial deficits (eg, A Cancellation time/Object Decision) precede those in memory (MMSE) was also found in refs. [40–42]. Our finding of early mathematical difficulty (GDA) agrees with ref. [42].

In Supplementary Material we demonstrated the superiority of our new event-based model over the classic event-based model which incorporates Gaussian mixture modeling (in simulation and in application to real data), especially for highly non-Gaussian data such as found in standard neuropsychological tests where ceiling effects and floor effects are common. When the input data are close to Gaussian, both models perform well and the classic event-based model is probably the better choice for model parsimony (see Supplementary Material). We note that our new model is capable of working “out of the box” on both Gaussian and non-Gaussian data, ameliorating the need for pre-processing steps traditionally used on non-Gaussian data, such as log transformation. However, it would be straightforward to add an extra model-selection step to choose the most appropriate mixture-model for each biomarker if desired.

We highlight some strengths and limitations of our study. First, a particular strength of our methodology, by design, is its flexibility to handle non-Gaussian data such as clinical test scores. This was verified by our extensive performance evaluation experiments in the Supplementary Material. Secondly, our neuropsychological battery intentionally tests cognitive domains specific to PCA and tAD, which is a strength in terms of both specificity and for comparing two dementia syndromes, but is a minor limitation in terms of discovery — that is, our data cannot discover whether an untested domain/function such as language or motor function is useful in these syndromes. As noted in many of the original publications on the method,<sup>8,9</sup> one limitation of the event-based model (in general, not our particular model) is the assumption of events being independent, which is not the case for cognitive tests which will often depend upon multiple domains to varying extent. Interpretation of results must keep this in mind.

Considering the rarity of PCA, our sample sizes are good, although modest. The event-based model is relatively robust to small sample sizes due to (1) its simplicity (the minimum viable sample size is simply the number of features); and (2) the model being designed to highlight uncertainty (through the positional variance diagram). In our experiments the minimum viable sample size was exceeded, often substantially (17 features, 20–94 observations per feature), resulting in the precision we observed. The effect of a small sample size on an



**FIGURE 3** Histogram of model stages assigned to healthy controls and patients with posterior cortical atrophy (PCA) and typical Alzheimer's disease (tAD)

event-based model is to increase uncertainty in the ordering of events, which would manifest as increased positional variance such as in the bootstrapped models in Figures 1B and 2B. A key advantage of the event-based model is that it explicitly highlights this uncertainty to reveal parts of the progression where the data may be insufficient to inform on the precise sequence.

Our results promote further clinical investigation of cognitive assessment in dementia, such as the early estimated abnormality of the A Cancellation task in PCA. This suggests that measures intended to assess executive function/attention but which feature a prominent visual search component (eg, Trailmaking tasks, Digit Symbol) are susceptible to visual processing deficits.

Our results motivate us to consider avenues for future work in both applications and methodological directions: characterizing cognitive decline in other neurodegenerative diseases (eg, Parkinson's dementia, Huntington's disease, familial Alzheimer's disease), including investigating the possibility of cognitive subtypes; application of such results in clinical trials; and working toward turning KDE mixture modeling into a more general statistical methodology. The first avenue will improve disease understanding, for which we foresee wider application of our new method on larger datasets involving more individuals and/or a more extensive battery of neuropsychological tests. Although the method has already been used in a few studies,<sup>26,43–45</sup> future work includes application to other multimodal datasets (eg, cognitive and imaging data) to reveal insights into the progression of neurological diseases. This will include embedding our method within the classic event-based model,<sup>8,9</sup> within the recently proposed discriminative event-based model,<sup>46,47</sup> and within data-driven disease subtyping algorithms such as Subtype and Stage Inference (SuStain).<sup>15</sup> The second avenue for future work turns this improved disease understanding into quantitative tools for clinical trials. This includes enrichment through precise stratification and subject selection, and using model stage as a data-driven composite cognitive end-point. This is ongoing work. The third avenue involves methodological advances, for which we envisage starting with performance characterization on more complex clus-

tering tasks – our formalism is quite general but our experiments involved only two components/clusters in the mixture model. Such investigations might also include application to feature normalization for supervised learning. Additionally, embedding KDE mixture modeling within a Bayesian framework would broaden its utility and enable the use of weakly informative priors on hyperparameters for constraining the clustering results. Such constraints can be important for ensuring realistic results in biological applications.<sup>48</sup>

In summary, our results verify clinical opinion on disease progression while revealing new insight into the fine-grained sequence of clinical decline in memory-led tAD and vision-led PCA. This fine-grained understanding promises clinical utility for informing earlier differential diagnosis, and prognosis through data-driven disease staging – both are relevant for clinical trials and patient management.

#### ACKNOWLEDGMENTS

We thank all participants in the PCA longitudinal study for donating their time to research. We also thank everyone at the Dementia Research Centre who helped with data collection over the course of twenty years, including Manja Lehmann, Tim Shakespeare, Aida Suarez Gonzalez, Amelia Carton, Ivanna Pavisic, and Dilek Ocal. The Dementia Research Centre at the UCL Institute of Neurology in London, UK is an Alzheimer's Research UK Coordinating Centre and is grateful for support from the Department of Health's NIHR Biomedical Research Centres funding scheme, the NIHR Queen Square Dementia Biomedical Research Unit, the MRC Dementias Platform UK, and the Leonard Wolfson Experimental Neurology Centre.

This work was supported by the UK Engineering Physical Sciences Research Council (EPSRC) through grant EP/M006093/1. Cognitive data collection was supported by grants to Sebastian J. Crutch from ESRC/NIHR (ES/L001810/1) and Alzheimer's Research UK (Senior Research Fellowship). Neil P. Oxtoby and Daniel C. Alexander acknowledge the EuroPOND project: This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 666992. Neil P. Oxtoby and Daniel

C. Alexander acknowledge support from The Michael J. Fox Foundation for Parkinson's Research, the Alzheimer's Association, Alzheimer's Research UK, and the Weston Brain Institute through "NetMON: Network Models Of Neurodegeneration" (BAND-15-368107, 11042). Neil P. Oxtoby and Daniel C. Alexander acknowledge support from the NIHR University College London Hospitals Biomedical Research Centre. Alexandra L. Young is supported by an EPSRC doctoral prize fellowship. Neil P. Oxtoby is a UKRI Future Leaders Fellow (MRC MR/S03546X/1).

## FUNDING

The funding agencies had no involvement in the research itself, including: study design; data collection, analysis, and interpretation; writing of this report; nor the decision to submit this article for publication.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## REFERENCES

- Cummings JL, Morstorf T, Zhong K. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. *Alzheimers Res Ther.* 2014;6(4):37.
- Cummings J, Aisen PS, Dubois B, et al. Drug development in Alzheimer's disease: the path to 2025. *Alzheimers Res Ther.* 2016;8:39.
- Cummings J, Lee G, Mortsdorf T, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2017. *Alzheimers Dementia.* 2017;3(3):367–384.
- Cummings J, Lee G, Ritter A, Zhong K. Alzheimer's disease drug development pipeline: 2018. *Alzheimers Dementia.* 2018;4:195–214.
- Cummings J, Lee G, Ritter A, Sabbagh M, Zhong K. Alzheimer's disease drug development pipeline: 2019. *Alzheimers Dementia.* 2019;5:272–293.
- Villemagne VL, Burnham S, Bourgeat P, et al. Amyloid  $\beta$  deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol.* 2013;12(4):357–367.
- Fonteijn HM, Clarkson MJ, Modat M, et al. An event-based disease progression model and its application to familial Alzheimer's disease. In: Székely G, Hahn HK, eds. *Information Processing in Medical Imaging*. Vol. 22. Berlin, Heidelberg: Springer; 2011:748–759.
- Fonteijn HM, Modat M, Clarkson MJ, et al. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage.* 2012;60(3):1880–1889.
- Young AL, Oxtoby NP, Daga P, et al. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain.* 2014;137(9):2564–2577.
- Oxtoby NP, Young AL, Cash DM, et al. Data-driven models of dominantly-inherited Alzheimer's disease progression. *Brain.* 2018;141(5):1529–1544.
- Chen G, Shu H, Chen G, et al. Staging Alzheimer's disease risk by sequencing brain function and structure, cerebrospinal fluid, and cognition biomarkers. *J Alzheimers Dis.* 2016;54(3):983–993.
- Wijeratne PA, Young AL, Oxtoby NP, et al. An image-based model of brain volume biomarker changes in Huntington's disease. *Ann Clin Transl Neurol.* 2018;5(5):570–582.
- Byrne LM, Rodrigues FB, Johnson EB, et al. Evaluation of mutant huntingtin and neurofilament proteins as potential markers in Huntington's disease. *Sci Transl Med.* 2018;10(458):eaat7108.
- Eshaghi A, Marinescu RV, Young AL, et al. Progression of regional grey matter atrophy in multiple sclerosis. *Brain.* 2018;141(6):1665–1677.
- Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat Commun.* 2018;9(1):221.
- Oxtoby NP, Alexander DC; for the EuroPOND Consortium. Imaging plus X: multimodal models of neurodegenerative disease. *Curr Opin Neurol Neurosurg.* 2017;30(4):371–379.
- Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimers Dement.* 2016;12(3):292–323.
- Amieva H, Mokri H, Le Goff M, et al. Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: a study of 20 years of cognitive decline. *Brain.* 2014;137(4):1167–1175.
- Machulda MM, Hagen CE, Wiste HJ, et al. Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status. *Clin Neuropsychol.* 2017;31(1):99–117.
- Dubois B, Feldman HH, Jacova C, et al. Advancing research diagnostic criteria for Alzheimer's disease: the iwG-2 criteria. *Lancet Neurol.* 2014;13(6):614–629.
- McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 2011;7(3):263–269.
- Warren JD, Fletcher PD, Golden HL. The paradox of syndromic diversity in Alzheimer disease. *Nat Rev Neurosci.* 2012;8(8):451–464.
- Crutch SJ, Schott JM, Rabinovici GD, et al. Consensus classification of posterior cortical atrophy. *Alzheimers Dement.* 2017;13(8):870–884.
- Tang-Wai DF, Graff-Radford NR, Boeve BF, et al. Clinical, genetic, and neuropathologic characteristics of posterior cortical atrophy. *Neurology.* 2004;63(7):1168–1174.
- Hof PR, Vogt BA, Bouras C, Morrison JH. Atypical form of Alzheimer's disease with prominent posterior cortical atrophy: a review of lesion distribution and circuit disconnection in cortical visual pathways. *Vision Res.* 1997;37(24):3609–3625.
- Firth NC, Primativo S, Marinescu R-V, et al. Longitudinal neuroanatomical and cognitive progression of posterior cortical atrophy. *Brain.* 2019;142(7):2082–2095.
- Mendez MF, Ghajarian M, Perryman KM. Posterior cortical atrophy: clinical characteristics and differences compared to Alzheimer's disease. *Dement Geriatr Cogn Disord.* 2002;14(1):33–40.
- Dubois B, Feldman HH, Jacova C, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* 2010;9(11):1118–1127.
- Dubois B, Feldman HH, Jacova C, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* 2007;6(8):734–746.
- Gorno-Tempini M, Hillis A, Weintraub S, et al. Classification of primary progressive aphasia and its variants. *Neurology.* 2011;76(11):1006–1014.
- Lehmann M, Barnes J, Ridgway GR, et al. Basic visual function and cortical thickness patterns in posterior cortical atrophy. *Cereb Cortex.* 2011;21(9):2122–2132.
- Yong KXX, Shakespeare TJ, Cash D, et al. Prominent effects and neural correlates of visual crowding in a neurodegenerative disease population. *Brain.* 2014;137(12):3284–3299.
- Yong KX, Shakespeare TJ, Cash D, Henley SM, Warren JD, Crutch SJ. (Con)text-specific effects of visual dysfunction on reading in posterior cortical atrophy. *Cortex.* 2014;57:92–106.
- Primativo S, Yong KX, Shakespeare TJ, Crutch SJ. The oral spelling profile of posterior cortical atrophy and the nature of the graphemic representation. *Neuropsychologia.* 2017;94:61–74.
- Groot C, Yeo BT, Vogel JW, et al. Data-driven detection of latent atrophy factors related to phenotypical variants of posterior cortical atrophy, bioRxiv 679225. <https://doi.org/10.1101/679225>.



36. Ridha BH, Crutch S, Cutler D, et al. A double-blind placebo-controlled cross-over clinical trial of DONepezil in posterior cortical atrophy due to underlying Alzheimer's Disease: DONIPAD study. *Alzheimers Res Ther.* 2018;10(1):44.
37. Suarez-Gonzalez A, Ocal D, Pavisic I, et al. ReadClear: an assistive reading tool for people living with posterior cortical atrophy. *J Alzheimers Dis.* 2019;71(4):1285–1295.
38. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2011;12:2825–2830.
39. Scott DW. On optimal and data-based histograms. *Biometrika.* 1979;66:605–610.
40. Lehmann M, Barnes J, Ridgway GR, et al. Global gray matter changes in posterior cortical atrophy: a serial imaging study. *Alzheimers Dement.* 2012;8(6):502–512.
41. Kennedy J, Lehmann M, Sokolska MJ, et al. Visualizing the emergence of posterior cortical atrophy. *Neurocase.* 2012;18(3): 248–257.
42. Ossenkoppele R, Cohn-Sheehy BI, La Joie R, et al. Atrophy patterns in early clinical stages across distinct phenotypes of Alzheimer's disease. *Hum Brain Mapp.* 2015;36(11):4421–4437.
43. Oxtoby NP, Garbarino S, Firth NC, Warren JD, Schott JM, Alexander DC; for the Alzheimer's Disease Neuroimaging Initiative. Data-driven sequence of changes to anatomical brain connectivity in sporadic Alzheimer's disease. *Front Neurol.* 2017;8:580.
44. Firth NC, Startin CM, Hithersay R, et al. Sequence of cognitive changes associated with development of Alzheimer's disease in Down syndrome – data driven analysis, bioRxiv 2018;263095. <https://doi.org/10.1101/263095>.
45. O'Connor A, Oxtoby NP, Weston PSJ, et al. The sequence and timing of preclinical cognitive decline in autosomal dominant Alzheimer's disease. *Alzheimers Dement.* 2019;15(7):882.
46. Venkatraghavan V, Bron EE, Niessen WJ, Klein S. A discriminative event based model for Alzheimer's disease progression modeling. In: Information Processing in Medical Imaging, Lecture Notes in Computer Science. Cham: Springer; 2017:121–133. [https://doi.org/10.1007/978-3-319-59050-9\\_10](https://doi.org/10.1007/978-3-319-59050-9_10).
47. Venkatraghavan V, Bron EE, Niessen WJ, Klein S. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. *NeuroImage.* 2019;186:518–532.
48. Lorenzi M, Filippone M. Constraining the dynamics of deep probabilistic models, arXiv (2018). arXiv:1802.05680v2.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Firth NC, Primativo S, Brotherhood E, et al. Sequences of cognitive decline in typical Alzheimer's disease and posterior cortical atrophy estimated using a novel event-based model of disease progression. *Alzheimer's Dement.* 2020;16:965–973. <https://doi.org/10.1002/alz.12083>