NeuroImage 278 (2023) 120279

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Disease progression modelling of Alzheimer's disease using probabilistic principal components analysis

Martin Saint-Jalmes [*,a,b,g], Victor Fedyashov [a,b], Daniel Beck [a,c], Timothy Baldwin [a,c,d], Noel G. Faux [a,b,e], Pierrick Bourgeat [f], Jurgen Fripp [f], Colin L. Masters [g], Benjamin Goudey [a,b], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] *ARC Training Centre in Cognitive Computing for Medical Technologies, University of Melbourne, Carlton, VIC, Australia*
[b] *The Florey Department of Neuroscience and Mental Health, The University of Melbourne, Australia*
[c] *School of Computing and Information Systems, The University of Melbourne, Australia*
[d] *Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates*
[e] *Melbourne Data Analytics Platform, The University of Melbourne, Australia*
[f] *CSIRO Health and Biosecurity, Brisbane, Australia*
[g] *The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Australia*

ARTICLE INFO

Keywords:
Alzheimer's disease
Disease progression modeling
Latent disease time
Principal components analysis
Machine learning

ABSTRACT

The recent biological redefinition of Alzheimer's Disease (AD) has spurred the development of statistical models that relate changes in biomarkers with neurodegeneration and worsening condition linked to AD. The ability to measure such changes may facilitate earlier diagnoses for affected individuals and help in monitoring the evolution of their condition. Amongst such statistical tools, disease progression models (DPMs) are quantitative, data-driven methods that specifically attempt to describe the temporal dynamics of biomarkers relevant to AD. Due to the heterogeneous nature of this disease, with patients of similar age experiencing different AD-related changes, a challenge facing longitudinal mixed-effects-based DPMs is the estimation of patient-realigning time-shifts. These time-shifts are indispensable for meaningful biomarker modelling, but may impact fitting time or vary with missing data in jointly estimated models. In this work, we estimate an individual's progression through Alzheimer's disease by combining multiple biomarkers into a single value using a probabilistic formulation of principal components analysis. Our results show that this variable, which summarises AD through observable biomarkers, is remarkably similar to jointly estimated time-shifts when we compute our scores for the baseline visit, on cross-sectional data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Reproducing the expected properties of clinical datasets, we confirm that estimated scores are robust to missing data or unavailable biomarkers. In addition to cross-sectional insights, we can model the latent variable as an individual progression score by repeating estimations at follow-up examinations and refining long-term estimates as more data is gathered, which would be ideal in a clinical setting. Finally, we verify that our score can be used as a pseudo-temporal scale instead of age to ignore some patient heterogeneity in cohort data and highlight the general trend in expected biomarker evolution in affected individuals.

1. Introduction

Alzheimer's Disease (AD) is a neurodegenerative condition which is projected to affect more than 100 million by 2050 (Brookmeyer et al., 2007), or around 150 million individuals for all types of dementia (Nichols et al., 2022). Recent biological redefinitions of the disease (Jack et al., 2018) have supported describing the underlying condition of AD patients from changes in associated biomarkers. Information

* Corresponding author at: ARC Training Centre in Cognitive Computing for Medical Technologies, Level 4, Melbourne Connect (Bldg 290), 700 Swanston Street, Carlton VIC 3010, Australia.

E-mail address: m.saintjalmes@unimelb.edu.au (M. Saint-Jalmes).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

https://doi.org/10.1016/j.neuroimage.2023.120279
Received 28 March 2023; Received in revised form 27 June 2023; Accepted 12 July 2023
Available online 15 July 2023

gathered from biomarkers may be useful in selecting individuals for clinical trials (Mattsson et al., 2015), or introducing lifestyle changes in patients with an early diagnosis (Scheltens et al., 2021). In clinical practice, providing individuals with concise information about their state and potential projections over time may also be helpful.

A description of the disease, and of a patient's state, may be performed in several ways. A trimodal staging system (Sperling et al., 2011) can be used by clinicians to provide a diagnosis establishing AD dementia, ruling individuals as cognitively normal (CN), or giving an intermediary status of mild cognitive impairment (MCI). Previous literature has explored predicting the correct diagnosis from biomarkers (Kumar et al., 2021; Wang et al., 2019). Alternatively, an individual's condition may be partly depicted through the lens of a single cognitive test (Fisher et al., 2019; Tabarestani et al., 2018), though such approaches may be limited by the tests' varying sensitivities at different stages of the disease (Abuhmed et al., 2021; Yagi et al., 2019). *Disease progression models* (DPM) form a third class of methods that were inspired by the dynamical model proposed by Jack et al. (2010), where a hypothesised clinical deterioration could be observed from successive degradations in biomarker measurements. DPMs are quantitative, data-driven models that aim to explain the longitudinal evolutions of biomarker trajectories and how changes relate to the disease (Donohue et al., 2014; Fonteijn et al., 2012; Guerrero et al., 2016; Koval et al., 2021; Kühnel et al., 2021; Li et al., 2019; Lorenzi et al., 2019; Mehdipour Ghazi et al., 2021; Schiratti et al., 2015; Venkatraghavan et al., 2019; Young et al., 2014).

Among DPMs, some of the most recent works are adaptations of longitudinal mixed-effects models, which are well suited to represent changes of biomarkers over time, with individual variations that are likely to emerge in a population. One of the specific challenges facing such models is the choice of a meaningful time scale. Neither time since study enrolment nor age are particularly appropriate scales that can reflect biomarker changes given the heterogeneity of AD conditions (Li et al., 2019). To address this issue, previous literature has proposed some form of temporal reparameterisation of age (Bilgel et al., 2017), or study time (Schiratti et al., 2015). Most often, this transformation is limited to the addition of a patient-specific time-shift, as a single value that allows for some degree of temporal realignment across study participants (Donohue et al., 2014; Kühnel et al., 2021; Li et al., 2019; Lorenzi et al., 2019).

We believe such time-shifts ultimately reflect the existing disparity in AD progression, as they attempt to remove temporal lags between patients to reach comparable disease states to fit mixed-effects models. As such, we propose a method that builds a single value from a varied range of biomarkers previously related to AD to reproduce the desired effect of these time-shifts. We compute estimates of this single value using the probabilistic formulation of principal components analysis (PPCA, Tipping and Bishop, 1999) applied to biomarker data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). We specifically leverage the previously daunting heterogeneity of the cohort to capture, using PPCA, the variability of observed biomarker values from enrolled β-amyloid positive patients at diverse points on the AD continuum, and summarise this heterogeneity in the first principal component. In addition, our proposed approach can be extended to utilise the repeat measurements of longitudinal data as they become available when patients undergo follow-up examinations. The value can therefore generalise as a continuous disease state score, indicative of AD progression. This approach is not unlike prior work from Kühnel et al. (2021), translating time-shifts into "predicted AD months" at each visit.

Our contribution is a robust approach that can separate, and thus temporally realign, heterogeneous patients, according to biomarkers collected on cross-sectional data. The resulting value can generalise to a progression score in the longitudinal setting, also highlighting changes in biomarkers. We are not aware of any prior literature employing PCA or its probabilistic formulation to describe disease states or temporal realignment in AD. PCA has otherwise been used for classification (Alam et al., 2017; López et al., 2009) and the first principal component was shown to change a diagnosis in latent space manipulation (Ayvaz and Baytas, 2021). In this article, we further describe the PPCA method used for the computation of our score, as well as the ADNI data and biomarkers involved in Section 2. In this same section, we detail the experiments we designed to validate the use of our method, including ensuring its robustness to missing data or biomarkers (Section 2.4). A presentation of the results is given in Section 3, before discussing their implications and limits (Section 4).

2. Material and methods

2.1. The Alzheimer's disease neuroimaging initiative (ADNI)

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether biological markers including magnetic resonance imaging (MRI), positron emission tomography (PET), cerebrospinal fluid (CSF) analyses, as well as clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease. A more detailed list of public and private parties involved in the ADNI project is given in the Acknowledgements section, and up-to-date information can be found on www.adni-info.org.

The collected data includes 15,789 records from 2383 unique patients, both new and rollover as of the latest ADNI-3 phase. Of these, 1022 patients who were amyloid-positive at baseline have been selected, corresponding to individuals on the AD continuum (Jack et al., 2018). These records were retrieved from the `ADNIMERGE` R package[2] (built February 22, 2022, most recent examination dated February 8, 2022). From the provided biomarkers, we make use of a range of established biomarkers covering amyloid, tau, neurodegeneration and cognitive tests. Table 1 presents some demographic information, as well as relevant biomarker statistics at the baseline visit for these individuals. Further information about biomarkers, including a description of these amyloid, tau, neurodegeneration and cognitive biomarkers used for our model, can be found in the supplementary material, Section A.1.

2.2. Data pre-processing and splitting

All experiments, aside from those in Section 3.7 and the supplementary material, focus on amyloid-positive individuals that can be considered to be on the AD continuum. This filtering is performed according to reported CSF concentrations, and is described in more detail in the supplementary material, Section A.2.

Volumetric biomarkers derived from MRI data have been adjusted for age and intracranial volume (ICV) at the baseline visit, as suggested by prior literature (Barnes et al., 2010). We specifically adjust these biomarkers as they showed the strongest correlations with age in cognitively normal, amyloid-negative individuals (Table A.5). As a result, the adjustment is performed on the basis of a linear model fit on these same cognitively normal and amyloid-negative individuals to address the effects of normal ageing on volumetrics. Non-volumetric biomarkers have not been adjusted for age, as prior analyses depicted significant, but weaker correlations (Table A.5). Further details of this analysis can be found in the supplementary material, Section A.5.

All biomarkers are scaled using z-score normalisation to reduce differences in scales and orders of magnitude (as shown in Table 1). For ease of interpretability, the sign of normalised biomarkers that typically decrease with age or cognitive impairment, such as volumes or fludeoxyglucose-PET, was reversed. As such, increasing biomarker

[2] https://adni.bitbucket.io.

Table 1
Baseline demographic information of β-amyloid positive patients, by baseline diagnosis. Unless otherwise specified (category: unknown), values represented here exclude missing data. Cells including percentages should be read as a count, followed by its percentage within each diagnosis group. Others depict the average unnormalised biomarker value followed by its standard deviation. Biomarkers in the bottom half of the table are used to fit the model, except in Section 3.7, where cognitive variables are excluded. All patients at baseline have at least five available biomarkers collected.

	CN	MCI	Dementia
N	260	510	251
Age	73.5 ± 6.2	73.0 ± 7.2	74.1 ± 8.3
Sex: Female	146 (56%)	206 (40%)	102 (41%)
Protocol: ADNI1	57 (22%)	156 (31%)	97 (39%)
Protocol: ADNIGO	0 (0%)	65 (13%)	0 (0%)
Protocol: ADNI2	120 (46%)	230 (45%)	122 (49%)
Protocol: ADNI3	83 (32%)	59 (12%)	32 (13%)
APOE ε4 (0)	144 (55%)	189 (37%)	68 (27%)
APOE ε4 (1)	96 (37%)	231 (45%)	125 (50%)
APOE ε4 (2)	17 (7%)	80 (16%)	56 (22%)
APOE ε4 (unknown)	3 (1%)	10 (2%)	2 (1%)
Years of education	16.4 ± 2.4	16.1 ± 2.8	15.5 ± 2.9
Hippocampus	7454 ± 880	6,717.8 ± 1075	5860 ± 1,031
Ventricles	36,832 ± 20,274	42,211 ± 23,284	50,191 ± 24,076
Whole Brain	1,039,478 ± 110,367	1,036,975 ± 109,244	995,104 ± 117,637
Entorhinal	3932 ± 623	3511 ± 738	2885 ± 747
FDG	1.3 ± 0.1	1.2 ± 0.1	1.1 ± 0.1
AV45	1.2 ± 0.2	1.3 ± 0.2	1.4 ± 0.2
RAVLT learning	6.0 ± 2.5	3.9 ± 2.6	1.9 ± 1.8
FAQ	0.2 ± 0.7	3.6 ± 4.1	13.0 ± 6.7
Adas-Cog-13	9.0 ± 4.4	17.4 ± 6.8	30.3 ± 8.2
Ptau 181	22.8 ± 11.4	31.2 ± 16.6	36.6 ± 15.5
Plasma NfL	36.7 ± 14.3	40.5 ± 17.9	47.2 ± 21.3
Total Tau	239 ± 104	311 ± 147	364 ± 141

values are likely indicative of decline.

The set of amyloid-positive patients is further split at random into two groups of equal size, hereafter named "train" and "test". This split is first used here for normalisation purposes, with the test set being rescaled according to the train set's biomarker values distribution. The split is also relevant to Section 3.1 and Section 3.4 in providing unbiased estimates of our constructed score estimate.

In the remainder of this article, all analyses are conducted on patients with at least four (*i.e.* one-third of) input biomarkers available. While there are no methodological limitations to selecting a lower threshold, this allows some degree of quality control, as it avoids imputing scores from a limited number of biomarker values and possible bias that may result from this. This filtering has no incidence on cross-sectional analysis, as all individuals meet this threshold on baseline data. However, only 918 of the 1022 enrolled had sufficient data in the follow-up visits.

2.3. Probabilistic principal components analysis (PPCA)

In order to capture the disparity of disease states present in the ADNI dataset, we use principal components analysis, a dimensionality reduction technique designed to build a projection to a transformed space while retaining maximal variability in the data.

Among the set of directions determined from the projection, the first principal component (PC1) is a subspace of a single dimension. Each observation can therefore be mapped to a single value in this subspace. Naturally, focusing on more principal components would allow for a better explanation of the variability of the original data, but limiting our approach to the first principal component ensures that we obtain an interpretable value, which we have compared to a progression score or to a time-shift.

We focus on the probabilistic formulation of principal components analysis outlined by Tipping and Bishop (1999), Bishop (2006). In the original article, the authors propose an expectation-maximisation (EM) procedure to iteratively build the latent space formed by the principal components, while considering an unknown source of noise in the observed data. This is particularly well suited to the context of biomarkers, which are expected to exhibit some level of noise from measurement methodology. As an illustration, Shaw et al. (2011) have reported an average inter-centre coefficient of variation (%CV) of 17.9% for β-amyloid measured from CSF.

Mathematically, the conditional distribution of observed data $\boldsymbol{x}$ (here, biomarkers) conditioned on the latent variable z (our underlying disease progression score) is given by

$$p(\boldsymbol{x}|z) = \mathscr{N}\left(\boldsymbol{x}\middle|\mathbf{W}z, \sigma^2\mathbf{I}\right), \tag{1}$$

where $\mathbf{W}$ is the projection of the latent variable into the original observed data space and σ controls the scale of the assumed isotropic Gaussian noise in the observed data.

While the original formulation includes a mean vector $\boldsymbol{\mu}$, it is not included here as our data was previously centred (Section 2.2).

To further determine the inverse projection of $\boldsymbol{x}$ to z, we need to inverse the covariance matrix $\mathbf{C} = \mathbf{W}\mathbf{W}^\top + \sigma^2\mathbf{I}$. An EM algorithm can then be used to alternatively compute the latent variable and update parameters $\mathbf{W}$ and σ. Porta et al. (2005) offer alternative EM update equations that specifically address the case of missing data.

In addition to its ability to consider noisy data, probabilistic PCA also offers an in-built mechanism for handling missing data which is desirable in the context of normalised biomarkers. A direct imputation of missing data, such as replacing values with population averages would be equivalent to a 0-imputation because of our z-score normalisation. In practice, we expect patients with an extreme underlying condition (either particularly healthy controls or patients already with dementia) to exhibit biomarker measurements far from the population average. In the case of PPCA, missing values are imputed using the data covariance matrix, supporting estimates of missing data that are more consistent with other observed biomarkers.

2.4. Experiments

2.4.1. Comparison with a label-aware projection

A natural question that arises from the use of an unsupervised method such as PCA is how its projection compares to similar methods that make use of labels, such as disease status in this case. We address this question by contrasting our method with linear discriminant analysis (LDA), whose projection is specifically guided by a classification target.

Specifically, for this experiment, we perform a one-dimensional projection of a 3-class linear discriminant analysis of our biomarker data. This method is label-aware, and is therefore the best linear projection to $\mathbb{R}$ in separating patients of different diagnoses, considering their biomarker values. We limit the projection to a one-dimensional one for ease of comparison with PPCA, and therefore only consider the first eigenvector associated to the largest eigenvalue of $\boldsymbol{\Sigma}^{-1}\mathbf{B}$. A more formal presentation of LDA is given in the supplementary material, Section A.6. This projection is compared against the scores we built using PPCA. As the goal is to assess the relevance of our PPC1 score, and not the quality of LDA, we deliberately overfit LDA on the test set with ground truth labels, thus obtaining an idealised diagnosis-informed projection. In contrast, our PPCA-derived scores for the same test set are as previously described, with EM iterations run on the train set. The studied data contains multiple missing observations, which is not naturally handled by LDA. We consider two simple cases of ignoring, and zero-imputing missing data.

2.4.2. Evaluation of robustness

The robustness of our proposed approach is evaluated in two ways. First, we wish to understand how computed scores may vary for a given patient if some biomarkers are systematically not collected.

Additionally, a second aspect we consider is robustness to data missing at random. Missing data is especially relevant in the context of datasets from clinical cohorts which often have missing measurements (Coley et al., 2011; Hardy et al., 2009).

We report the impact of such scenarios by comparing the difference, for each patient at their baseline visit, between the scores previously computed (from what is hereafter called the "reference" model) and new scores from models that were refitted without specific biomarkers or with randomly masked data. We study the distributions of these differences and compute mean absolute differences between these scores,

$$\overline{\delta} = \frac{1}{N}\sum_{i=1}^{N}\left|s_i - s_i^{'}\right|, \tag{2}$$

for all patients $1..N$, where s_i is the PPCA PC1 score for patient i with the "reference" model, and $s_i^{'}$ their score with the alternative model.

By design, PPCA is able to fit a transformation even with some missing observations, alternating between imputation and maximising the likelihood within the EM algorithm. We introduced a varying amount of masked data, to simulate missing measurements at random. For analysis, only records that still had at least four biomarker values (one-third of input features) after masking were retained, the same condition used for the other analyses. This is done as to limit bias that may arise from visits where only few, extreme measurements are collected.

2.4.3. Cut-off point-based diagnosis predictions

In order to gain some intuition of our computed scores, we shift the current setting to the previously mentioned task of disease classification. In particular, the goal is to evaluate the adequacy of our score to predict a patient's assigned diagnosis. This disease classification is not the main contribution of this paper and still carries the issues of label uncertainty and limited information due to its discrete nature described in Section 1 but this can give some interpretation of our scores.

As this paper does not seek to propose a disease classification model, but rather to validate this score, we select a model with minimal learnable parameters, namely decision trees.

In practical terms, we determine a threshold, or a cut-off using shallow decision trees (of depth one) to predict for each patient whether they should be assigned one of two diagnoses based on their computed PPCA PC1 score. This threshold is determined using the aforementioned train set, using the Gini index to measure node impurity on either side of the threshold. In this context, node impurity is measured as how mixed patients of either diagnoses are below or above the threshold. This approach is repeated for each pair of diagnoses. Impurity is defined for each node m and parameterised by a candidate threshold T as

$$Q_m(T) = \sum_{k=1}^{K}\widehat{p}_{mk}(1-\widehat{p}_{mk}), \tag{3}$$

where K is the number of possible classes, *i.e.* in our case, the number of possible diagnoses ($K = 2$, since we consider pairs of diagnoses), and $\widehat{p}_{mk}$ is the empirical proportion of observations of class k in node m. Intuitively, this approach best separates patients that are supposed to have different diagnoses, while relying on fewer parameters than a linear model.

2.4.4. Hierarchical model for longitudinal interpretation

Our score has so far been presented as a value placing patients of different underlying states on a one-dimensional scale, interpretable as the Alzheimer's continuum. As such, reported results have only considered cross-sectional data from the first ("baseline") visit at patients' enrolment. We further believe that there is no particular constraint in limiting this approach to baseline data. The score can be computed as a transformation of biomarker data for any visit of a patient, and can therefore be estimated for all follow-up visits.

We consider a hierarchical model to represent the longitudinal evolution of scores s computed at each visit. The goal is to determine, for each patient p, a pair of coefficients $[\alpha_p, \beta_p]$ describing their scores as a (noisy) affine transformation of time. The model is described in Eq. (4). As we suppose intercepts and slopes to be correlated, with patients in worse condition at baseline having faster further cognitive decline, we use a multivariate normal prior on the pairs of coefficients.

$$\begin{aligned}
s &\sim \mathcal{N}(\mu,\sigma)\\
\mu_i &= \alpha_p + \beta_p \cdot t_i\\
\begin{bmatrix}\alpha_p\\ \beta_p\end{bmatrix} &\sim \mathrm{MVNormal}\left(\begin{bmatrix}\alpha\\ \beta\end{bmatrix},\Sigma\right)\\
\alpha &\sim \mathcal{N}(0,5)\\
\beta &\sim \mathcal{N}(0,5)\\
\sigma &\sim \Gamma(1,1)\\
\Sigma &= \begin{bmatrix}\sigma_\alpha & 0\\ 0 & \sigma_\beta\end{bmatrix}\Omega\begin{bmatrix}\sigma_\alpha & 0\\ 0 & \sigma_\beta\end{bmatrix}\\
\sigma_\alpha &\sim \exp(1)\\
\sigma_\beta &\sim \exp(1)\\
\Omega &\sim \mathrm{LKJcorr}(2)
\end{aligned} \tag{4}$$

2.5. Comparison with other disease progression models

For relevant experiments presented above, we compare our model with three different methods, for which we compute time-shifts and scores on the same patients. The Multivariate Continuous-time Disease Progression model (MCDP, Kühnel et al., 2021) and the Latent Time Joint Mixed-effects Model (LTJMM, Li et al., 2019) are two examples of disease progression models, presented in the introduction (Section 1) that compute individual-specific time-shifts while describing biomarker evolution through mixed-effects models. Their time-shift parameters are estimated given longitudinal data, but are not time-varying, so they are compared against our PPC1 score computed cross-sectionally at the baseline visit. While our approach does not fully model the biomarker dynamics as in other DPMs, the goal of this analysis is to assess whether PPCA PC1 captures similar information to such time-shifts, with just baseline data.

The third method we compare is a plain average of input z-score normalised biomarkers in lieu of PPCA, at the baseline visit. Aside from MCDP, which only targets cognitive outcomes, the same biomarkers are utilised.

Further implementation details, including links and parameters used to fit these models, can be found in the supplementary material, Section A.9.

3. Results

3.1. Scoring patient visits

We compute scores based on the method described in Section 2.3. We first focus on the computed scores for baseline visits, pertaining to patients from the train and test sets. As discussed previously, the test set has been normalised according to the biomarker distributions from the train set. Additionally, scores for the test set at baseline are computed with the covariance matrix resulting from EM iterations on the train set.

Fig. 1 a) shows the resulting distribution of computed scores for these patients from the test set, corresponding to their baseline visit, grouped by diagnosis.

A good separation of patients can be observed, with the first principal component of PPCA capturing the heterogeneity of Alzheimer's Disease. As expected, diverse individuals enrolled in the ADNI cohort are linked to a range of different states, scored approximately in $[-5, 6]$, in the

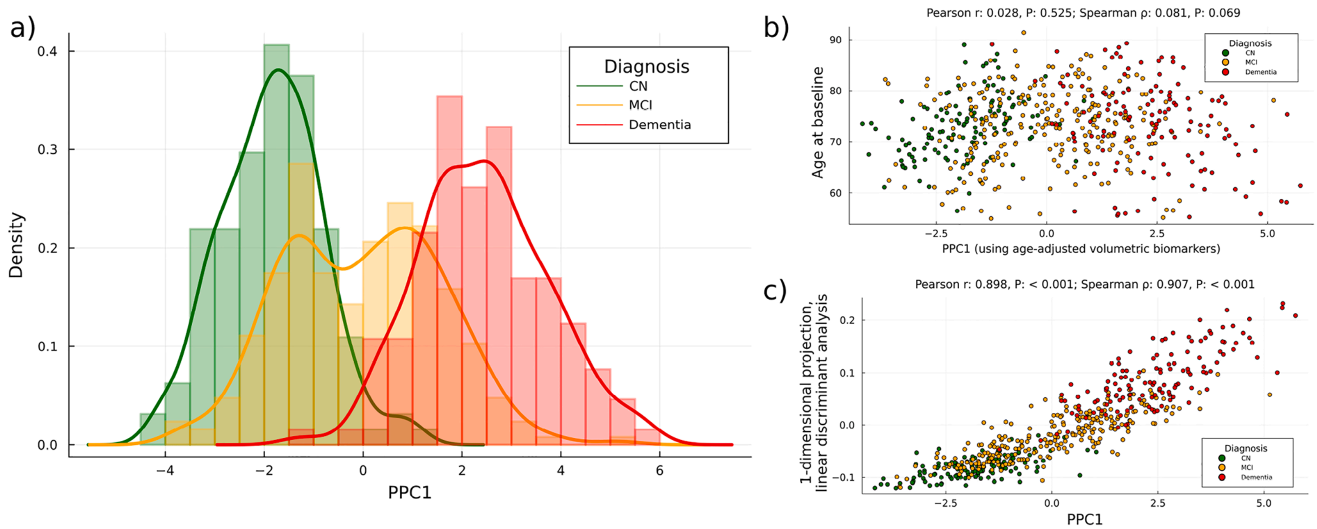

Fig. 1. Description of computed values using PPCA for 510 amyloid-positive patients from the test set, at their baseline visit: a) Per-diagnosis densities of calculated PPCA PC1s. Without leveraging diagnosis information, the first principal component captures the variability of the data and spreads apart patients of different diagnoses. b) Association of PPCA PC1 scores and age for patients of the test set, at the baseline visit. As opposed to Figure A.4, the scores were computed from adjusted volumetric biomarker using residuals from a linear model using age and intracranial volume at baseline, fitted on amyloid-negative, cognitively healthy individuals. c) Comparison of scores obtained through PPCA, and computed scores via a projection of a 1-dimensional linear discriminant analysis. For LDA, 0-imputation was performed when biomarkers were missing.

progression of Alzheimer's Disease.

Considering overall trends in diagnoses, which were not used for PPCA, this plot also offers a natural interpretation of our PPC1 score, overall placing healthier individuals on the left of the axis, and patients who have already experienced significant decline on the other end. This should be linked to the clinical disease stage plot proposed by Jack et al. (2010).

However, the lack of clear and complete separations between patients that are assigned different diagnoses illustrates some of the difficulties in relying solely on diagnosis as an outcome variable. In particular, while there is little overlap in scores from patients determined to be cognitively normal (CN) and those having dementia, patients with a mild cognitive impairment diagnosis are likely to have similar values for biomarkers (and therefore computed scores) to those with either other diagnosis.

We present in Fig. 1 b) the association of our score with age, and observe no significant correlation (Pearson $r = 0.028$, $P = 0.525$; Spearman rank $\rho = 0.081, P = 0.068$). In the collapse of biomarker information in a single value, the computed score therefore exhibits the variability in disease states, and is not simply reproducing trends linked to age.

3.2. Comparison with a label-aware projection

As described in Section 2.4.1, we chose to compare our PPCA-derived score with a single dimension, label-aware projection of input biomarkers using linear discriminant analysis (LDA). We report here the results for two comparisons, depending on the strategy chosen to handle missing data.

In Fig. 1 c), we replace missing data for biomarkers by 0, as a z-score normalisation was previously performed. This imputation approach biases missing biomarker measurements towards the mean, which may not accurately represent values that would have otherwise been observed for patients on more extreme sides of the Alzheimer's disease continuum. To address this, we also consider a similar experiment (Figure A.5) that was restricted to the 194 patients from the test set that had no missing biomarker values at baseline. A drawback of this approach is that the sample size for comparison is lower.

PPC1 scores obtained through our method are strongly correlated with the optimal one-dimensional linear projection determined by multiclass LDA, whether we zero-impute missing data ($r = 0.898, P < 0.001; \rho = 0.907, P < 0.001$) or when we only consider observations with all biomarkers available ($r = 0.904$, $P < 0.001; \rho = 0.898$, $P < 0.001$). Given that the LDA approach is the optimal one-dimensional, diagnosis-aware separation that can be achieved, these correlations constitute further evidence that our PPCA-based score provides a useful separation of patient disease status and may represent a sensible estimate of their underlying disease progression.

3.3. Robustness

We report the results of our robustness experiments in this section. The first of such experiments consists of removing some of the variables from the computation of our score. There are different situations in which this may be of interest. It is first important to exclude the possibility that our score may solely be determined from a single or a handful of biomarkers, with others providing little to no contribution to its computation. Alternatively, we want to ensure that our score is still meaningful when some biomarkers are not collected, due to the cost of equipment (*i.e.* in the absence of MRI-derived brain volume information), or due to procedure invasiveness (*i.e.* without any CSF measurements).

Fig. 2 describes the effects of removing one given variable on the scores of each individual. Regardless of the biomarker removed, the first and third quartiles encompassing the central 50% of these differences are within $[-0.30, 0.25]$ and $\overline{\delta} < 0.35$, which are small comparatively to the typical range of $[-5, 6]$ for our computed PPCA PC1 scores seen in Fig. 1 a). Similar results are observed when we removed pairs of variables (Supplementary material, Figure A.11).

We can attempt to interpret some of the more apparently extreme observations in Fig. 2. For example, one patient had a score difference $s_i - s_i^{'} \approx 3$ when removing total tau as a biomarker. A high difference suggests this specific patient had a much lower score ("trending less" towards dementia) when removing total tau. With tau being a marker of non-AD specific dementia, it may be that such patients have other dementias or concomitant suspected non-Alzheimer's pathologic change with dementia.

We also examine the impact of certain classes of biomarkers, for

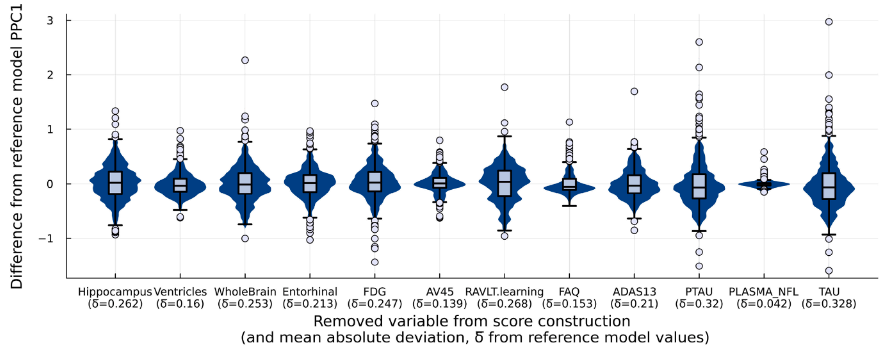

Fig. 2. Distributions of differences between PPCA PC1 scores of each individual from the test set (511 amyloid-positive patients) at their baseline visit, computed with all biomarkers available ("reference model") and models that excluded each biomarker from score construction.

example by removing cognitive test scores, which may otherwise be regarded as a source of concern regarding circularity with diagnosis. Figure A.6 shows that the scores are robust when removing CSF measures, volumetrics or cognitive measures. Results only begin to change substantially when only PET β-amyloid (AV45) and CSF phosphorylated-tau 181 are used as inputs to the score, with a much greater mean absolute difference $\overline{\delta} = 1.327$. This discrepancy likely highlights the biomarkers' differing temporal sensitivity, with our reference score computation taking into account cognitive biomarkers as well.

Finally, we also assess the model's robustness to data missing at random. Table 2 shows the mean absolute differences for increasing proportions of masked data. As our method imputes missing measurements from those available, it can be seen that our scores vary very little, with $\overline{\delta} \leq 0.4$ even when masking 30% of data. This degree of robustness highlights the suitability of our score to real-world settings where missing data is common.

3.4. Diagnosis prediction from cut-offs

In this section, we report the observed classification performance resulting from the thresholding procedure described in Section 2.4.3. We reproduce classification metrics in Table 3 for the test set, and with more details in the supplementary material, Table A.6. The thresholds of interest are determined on the basis of the train data only. Specifically, for the pairs of CN/Dementia, CN/MCI, and Dementia/MCI, the optimal thresholds are -0.06, -0.76 and 1.56, respectively. F_1-scores for each task were all above 0.75, and performance is comparable for the train and test sets. We note however that the CN/MCI task has lower performance, which is in line with current literature, as this is regarded as the harder task.

Table 2
Mean and standard deviation of absolute differences between PPCA PC1 scores of each individual from the test set with all available data used, and recomputed after a percentage of data was randomly masked. The left column reports scores when this masking procedure was applied only to the test data (*i.e.* with no impact on PPCA fit) and the right shows statistics of these differences when both train and test data were masked. Results averaged over 100 replications. Standard deviations across replications of the presented means were below $3{\cdot}10^{-2}$ in all situations.

	No train data drop	With train data drop
10%	0.16 ± 0.23	0.18 ± 0.22
20%	0.29 ± 0.31	0.30 ± 0.31
30%	0.39 ± 0.39	0.40 ± 0.39
40%	0.49 ± 0.47	0.51 ± 0.47
50%	0.60 ± 0.54	0.65 ± 0.57
60%	0.73 ± 0.64	1.16 ± 0.81

Table 3
Decision tree-based disease classification on test data, from thresholds determined on the training set (depth of one, using Gini index as a criterion, baseline visits only). These thresholds were applied to separate patients according to their computed PPCA PC1 score, turning the underlying disease time estimation into a diagnosis classification problem.

	Accuracy	F_1-score	Precision	Recall
CN *vs.* Dem.	0.973	0.973	0.962	0.985
CN *vs.* MCI	0.721	0.751	0.920	0.635
Dem. *vs.* MCI	0.806	0.851	0.865	0.837

3.5. From realigning time-shift to progression

We now consider the results of the longitudinal model presented in Section 2.4.4, where we tried to find the best description of patient score trajectories as a line. Posterior draws have been extracted after Markov chain Monte Carlo (MCMC) sampling for the model, using longitudinal data of patients from the test set.

Fig. 3 shows the affine progressions representing each patient. Grouping patients *a posteriori* based on the diagnosis they were given at their baseline visit, we also present average progressions for each category. As we previously hypothesised, patients already experiencing cognitive decline typically start with a higher score but also see it increase much faster than healthy individuals. Patients with a MCI diagnosis at baseline see much more heterogeneous trajectories, with some of them progressing as fast as those with dementia. While this is an expected outcome of the wide MCI diagnosis, it is also a limitation of grouping individuals on the basis of a single visit, with fast progressors likely being patients who are soon after given a dementia diagnosis. We verify this hypothesis in the supplementary material, Section A.8.

In the supplementary material, Figure A.3, we reproduce full

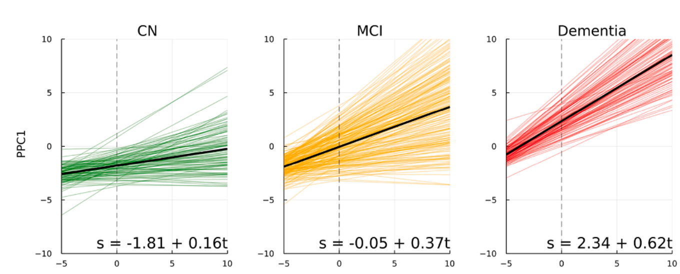

Fig. 3. Individual (coloured) and average (black) trajectories from posterior draws of slope and intercept parameters, as specified by the model in Eq. (4). 451 individuals from the test set with at least two visits are included. Trajectories are separated *a posteriori* according to the diagnosis assigned to each patient at their baseline visit. Lines correspond to the estimated affine specifications and may extend beyond observed data.

posterior distribution for our parameters, including those corresponding to patient intercepts and slopes.

3.6. Comparison with other disease progression models

In the following section, we briefly discuss our findings resulting from comparing our score against previous works in disease progression modelling (Section 2.5), and the limits of such comparisons.

We observe a particularly strong correlation between our PPCA PC1 scores at baseline, and the latent disease time-shifts from LTJMM ($r = 0.983, P < 0.001$; $\rho = 0.983, P < 0.001$, supplementary Figure A.8). For each patient, LTJMM's sampled posterior mean for δ corresponds to the optimal time-shift that best describes longitudinal biomarker evolutions within a mixed-effect model. A strong correlation can also be seen between our scores and MCDP's "predicted disease months" at baseline ($r = 0.813, P < 0.001$; $\rho = 0.834, P < 0.001$).

The same cutoff-based classification task described in Section 2.4.3 is applied to these time-shifts and scores, to measure their ability to distinguish patients of different baseline diagnoses, relatively to our own score. The same train/test split is used for this task in determining thresholds, and predicting diagnosis for individuals in the test set. In Fig. 4, we report the confusion matrices of predictions derived from PPCA PC1s, against each of the three methods. Unsurprisingly, LTJMM has the highest agreement with our score, in line with the previously observed high correlations between estimated values of the two methods. Average z-scores consistently predict worse diagnoses than PPCA, which may reflect the impact of higher z-scores otherwise under-weighted by our method for biomarkers such as ventricles and plasma NfL (Fig. 5). MCDP, which only uses MMSE and ADAS-13 as outcomes, predicts overall the same diagnoses as with PC1, but yields the highest number of disagreements of the three methods.

In addition to differing sets of biomarkers, there is a caveat to the use of MCDP in this task, as the model explicitly utilises baseline diagnosis to estimate a shift between patients, resulting in a favourable bias. Full results of this decision tree-based classification task for all methods are presented in the supplementary material, Table A.6. While PPCA does not always surpass other methods' predictive performance, it does not perform much lower than methods computing estimates requiring longitudinal data.

3.7. Patient realignment

The literature of mixed-effects disease progression models, including the aforementioned LTJMM (Li et al., 2019), often depicts individual time-shifts as a necessity for patient realignment, where study time or age are not ideal to model biomarker evolutions.

In Section 3.6, we have seen that our score computed from baseline visits may be very similar to estimated time-shifts. Therefore, we conduct an additional analysis to study our method's ability to realign patients for unseen biomarkers. We compute our score anew on both amyloid negative and positive patients, excluding all cognitive

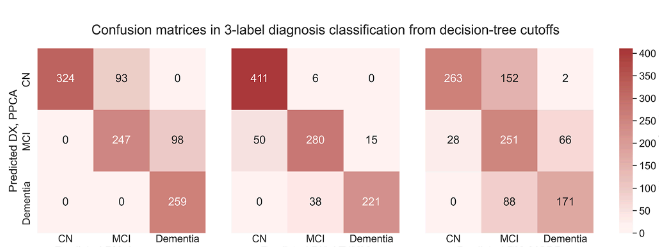

Fig. 4. Confusion matrices between PPCA PC1 and other methods' predictions in a 3-label diagnosis classification task from decision tree-determined thresholds. LTJMM: Latent Time Joint Mixed Effect Models (Li et al., 2019). MCDP: Multivariate Continuous-time Disease Progression model (Kühnel et al., 2021).

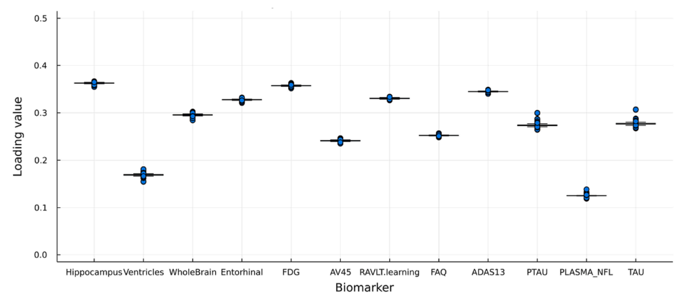

Fig. 5. Distribution of loadings determined from running PPCA EM iterations on the patients from the train set (511 amyloid-positive patients) at their baseline visit, across 100 random initialisations.

variables. Fig. 6 allows us to view the association between this unbiased score and a cognitive test, ADAS-13.

We observe a much higher correlation of ADAS-13 with our score – which was unbiased to cognitive performance – than with age. Although individual trajectories appear much more interpretable with age, this seems only especially true with individuals already living with dementia from the first visit, and there is otherwise only a mild correlation. In our case, lower PC1 individuals tend to have lower ADAS, and the use of PC1 as a "temporal" scale better highlights increases in ADAS than time measured by patient age.

4. Discussion

In this study, we have proposed using the first principal component of probabilistic principal components analysis to describe the underlying condition of patients biologically on the Alzheimer's continuum using a single value. The heterogeneity of underlying disease states in individuals enrolled in large Alzheimer's Disease cohorts such as ADNI offers a spectrum of measured biomarkers spanning a good range of the continuum and principal components analysis naturally provides a mechanism to quickly and robustly find a low-dimensionality representation of this variability. The choice of its probabilistic version emerges from the nature of biological data, in which some biomarker values may not always be available for all patients, and when they are available, they are inherently affected by measurement noise. Evaluations of our approach on amyloid positive individuals finds that it provides a strong separation of cross-sectional clinical diagnosis, with this unsupervised approach comparable to supervised models and DPMs requiring longitudinal data. We adapt our model to be able to estimate longitudinal rates of changes, allowing individual level trajectory predictions. Finally, we benchmark our approach against other disease progression models with more sophisticated underlying models and find strong correlations between all approaches, providing further justifications for our PPCA approach.

4.1. Linearity and disease progression models complexity

Disease progression score methodologies have tended to grow in complexity over time. From specific assumptions on the explicit shape of biomarker trajectories (Donohue et al., 2014; Jedynak et al., 2012), authors have recently explored models where functional forms are inferred from data (Lorenzi et al., 2019; Rudovic et al., 2019). Additionally, there is a trend towards greater flexibility in measuring patient-specific deviations from population-level models, whether through individual time-shifts, acceleration factors (Koval et al., 2021), or biomarker orderings (Venkatraghavan et al., 2019). This additional complexity comes at a cost, as a larger set of parameters must be fit to the data, increasing the risk of over-fitting and making the models slower and less robust. Our approach explores the impact of reducing complexity, assuming that the variability in a given dataset can be

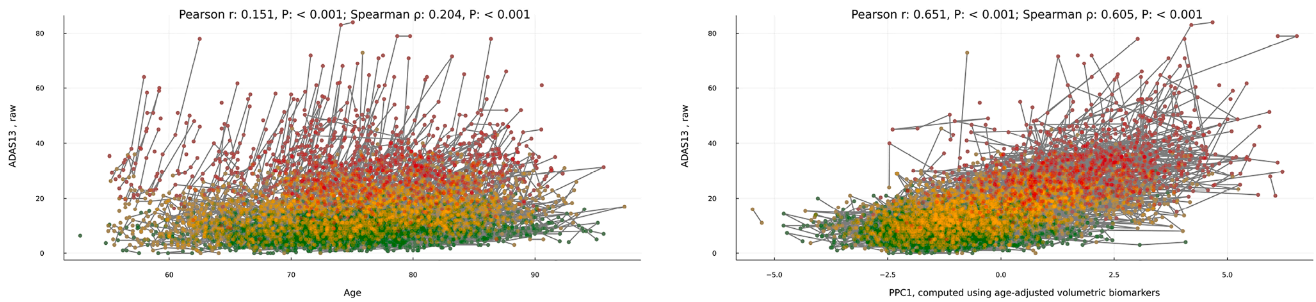

Fig. 6. Individual ADAS13 evolutions with respect to age (left) and PPCA PC1 score (right). In this instance, the score has been computed by excluding all cognitive variables, and on a superset of amyloid positive and negative patients. Visits are coloured by their patients' diagnosis at baseline. 6495 visits of 1738 test set individuals are considered.

modeled as a linear combination of available biomarkers. Surprisingly, our approach builds subject summaries that are comparable to those established in more complex, recently published disease progression models. While assumptions of longitudinal linearity may not always hold true biologically, parsimonious linear models may also be useful in informing patient-level trends over the long course of the disease. Our approach, which computes a score by normalising biomarkers and determining and optimal covariance matrix through EM iterations can be run in less than a minute on modern hardware for typical Alzheimer's datasets. Other disease progression models from the literature have, in our experience, been slower to fit or may have suffered from convergence issues.

4.2. Interpreting biomarker contributions

Our approach naturally offers a simple means to verify that our fit is sensible. We can extract the loadings matrix, which is in our case one-dimensional, and consider the value determined for each biomarker. This value is effectively a weight or a contribution to the PPCA PC1 score and allows us to understand how the resulting scores are computed. As seen in Fig. 5, there is no issue linked to scores being entirely determined by a few biomarkers. Conversely, we do not observe extremely low loading values for any one biomarker, suggesting the use of a group of biomarkers rather than a select few would be useful in scoring patients. As we previously normalised our data and harmonised the direction of abnormality (Section 2.2), scale differences or negative cancellation effects are not an issue here. Experiments described in Section 2.4.2 further support that our method is still applicable with comparable results in the absence of some of the biomarkers used, or with more missing data. This includes the exclusion of cognitive biomarkers, as presented in robustness experiments (Figure A.3) and Section 3.7, where we fit a model blind to cognitive measures.

4.3. Correlation of our score with age

As we have summarised disease states and progression in a single value per visit and patient, one notable risk would be that the computed score ultimately captured the effects of age, *i.e.* an increase correlated with normal ageing. Prior literature in disease progression modelling has also been concerned with this aspect in their models, although different approaches have been studied. Li et al. (2019), Raket (2020) have included age as a covariate within mixed-effects models, while Lorenzi et al. (2019) examined whether their time-shift was correlated with age.

In light of high correlations with age of the MRI volumetric biomarkers even in healthy amyloid-negative individuals (up to $r = 0.32$, supplementary material, Table A.5), MRI biomarkers were adjusted for intracranial volume and age. The adjusted MRI biomarkers effectively capture atrophy rather than normal ageing or heterogeneity in head sizes and lead to a progression score that has no significant correlation with age (Fig. 1 b). Conversely, computing scores with unadjusted volumetrics shows a slight, but significant correlation with age (supplementary material, Figure A.4). However, there may be implications to performing this adjustment. Voevodskaya et al. (2014) described some potential issues linked to loss of absolute information when adjusting for intracranial volume, using a similar regression approach.

4.4. Biomarker-time constructions and disease progression models

Disease progression models, especially those concerned with estimating subject-specific time-shifts, share some common goals with broader "biomarker-time" approaches. In the context of AD, there is growing interest in "amyloid-time" constructions (Betthauser et al., 2021; Schindler et al., 2021) to describe patients on a common scale and inform biomarker changes. Similar works have explored synchronising and following patients through cognitive tests (Yang et al., 2011) or studying potentially disease-related atrophy with imaging (Smith et al., 2019). A common theme for these disjoint areas of research is linked to the relative analysis of individuals that may not otherwise be comparable on a natural time-scale. Addressing this issue of heterogeneity with the choice of a biomarker time-scale, or estimating patient time-shifts in DPMs, are two ways enabling further study of progression patters. Relating our PPCA score to have a basis in terms of time (comparable to amyloid-time) is one avenue of future research.

4.5. Limitations & future works

We acknowledge here some of the limitations of the proposed approach and present some opportunities for future research directions. First, as we employ probabilistic principal components analysis for the projection of our biomarkers into scores, our method is currently limited to quantitative biomarkers. In practice, it would be sensible to include other demographic information such as sex or *APOE* ε4 status.

Our approach is centred around the construction of a single value that is meant to be representative of patients' underlying AD condition. The restriction of selecting only the first principal component allows us to have this degree of interpretability, but also masks other sources of variability in biomarkers. As we constrain an alignment on a "typical" AD scale, this model does not account for potential subtypes or different presentations of AD (Young et al., 2018).

With interpretability in mind, and as our approach allows the inspection of loadings to that end, we believe there it would be interesting to study a finer granularity for biomarkers, especially considering recent work in domain-level sub-scores for cognitive test results (Shan et al., 2021). Similarly, finer-grained analyses can be conducted on the longitudinal interpretation of our score, using clustering for example to identify groups of individuals that will develop dementia, as opposed to those who will only decline from normal ageing.

While we described the potential use of a score for realignment, the use of PPCA PC1 as a pseudo-time scale requires further work in its interpretability, as it is not quantifiable in natural time. If biomarker evolution can only be described in shifted, transformed, or pseudo-time scales, communication of progression scores remains challenging. Putting computed scores in perspective with projections with linear trends (Section 3.5) can be a means to describe whether an individual is progressing faster or slower than expected at a new visit.

The current use of PPCA over PCA is the ability to account for missing data and for dealing with noisy measurements. However, there are several opportunities to expand our handling of uncertainty in the PCA score in the subsequent model. By design, PPCA models the latent variable as Gaussian, with unit variance. Thus one approach is to a fully Bayesian PCA model (Bishop, 1998) which could provide us with posterior distributions on the noise of individual scores obtained from the latent dimension. With regards to our classification work, we are aware that prior works, such as that of Tami et al. (2018), have explored the use of uncertainty information in the context of decision trees, with splitting rules based on probability distributions, which could enable a future avenue for deriving probabilistic classifications. Finally, uncertainty estimates could also be used in our longitudinal model (Equation 4) to refine our priors for score noise when estimating individual trajectories.

Finally, we acknowledge that the current method has only been applied to data from the Alzheimer's Disease Neuroimaging Initiative. As a result, we are limited to patients satisfying ADNI's inclusion criteria, with a particular set of demographics that may not be similar to those of other cohort studies in Alzheimer's Disease research. While we believe our approach can be reproduced on other datasets, further validation is required.

5. Conclusion

We have proposed a method making use of the first principal component of probabilistic PCA to describe the underlying state of individuals on the Alzheimer's continuum as a single score from a set of relevant biomarkers. This score, which is easily re-computed for new patients, can be interpreted and used in different ways: it can provide intuition about disease states from the first visit, as variability captured from biomarkers reflects differences clinical diagnoses. It may be useful in realigning individuals for mixed-effects models, as it closely matches with patient time-shifts of other disease progression models. It may show how specific biomarkers evolve as scores worsen and finally it can be interpreted longitudinally as data for multiple visits become available over time. Further work will involve validation in other cohort studies.

Data availability

The data used in this work is available upon application to the Alzheimer's Disease Neuroimaging Initiative (ADNI, adni.loni.usc.edu) and acceptance of its data use agreement. Informed consent was obtained from enrolled participants by ADNI. The Probabilistic Principal Components Analysis algorithm is a re-implementation in Julia, derived from Jakob Verbeek's Matlab script made available for non-commercial use at lear.inrialpes.fr/~verbeek/software.php. Further analysis scripts are available upon request.

Declaration of Competing Interest

The authors declare that they have no competing interest.

Acknowledgments

This research was supported by the Australian Research Council Training Centre in Cognitive Computing for Medical Technologies (project number IC170100030) and funded by the Australian Government.

The Florey Institute of Neuroscience and Mental Health acknowledges the strong support from the Victorian Government and in particular the funding from the Operational Infrastructure Support Grant.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.;Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.;Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.neuroimage.2023.120279.

References

Abuhmed, T., El-Sappagh, S., Alonso, J.M., 2021. Robust hybrid deep learning models for Alzheimer's progression detection. Knowl. Based Syst. 213, 106688. https://doi.org/10.1016/j.knosys.2020.106688.

Alam, S., Kwon, G.R., Initiative, T.A.D.N., 2017. Alzheimer disease classification using KPCA, LDA, and multi-kernel learning SVM. Int. J. Imaging Syst. Technol. 27, 133–143. https://doi.org/10.1002/ima.22217.

Ayvaz, D. S., Baytas, I. M., 2021. Investigating conversion from mild cognitive impairment to Alzheimer's disease using latent space manipulation. arXiv:2111.08794 [cs].

Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., Fox, N.C., 2010. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 53, 1244–1255. https://doi.org/10.1016/j.neuroimage.2010.06.025.

Betthauser, T.J., Koscik, R.L., Jonaitis, E.M., Van Hulle, C.A., Basche, K.E., Kohli, A., Suridjan, I., Kollmorgen, G., Chin, N.A., Mueller, K.D., Clark, L.R., Christian, B.T., Okonkwo, O.C., Bendlin, B.B., Asthana, S., Carlsson, C.M., Zetterberg, H., Blennow, K., Johnson, S.C., 2021. Amyloid time: quantifying the onset of abnormal biomarkers and cognitive impairment along the Alzheimer's disease continuum. Alzheimer's Dement. 17 https://doi.org/10.1002/alz.056269.

Bilgel, M., Koscik, R.L., An, Y., Prince, J.L., Resnick, S.M., Johnson, S.C., Jedynak, B.M., 2017. Temporal order of Alzheimer's disease-related cognitive marker changes in BLSA and WRAP longitudinal studies. J. Alzheimers Dis. 59, 1335–1347. https://doi.org/10.3233/JAD-170448.

Bishop, C., 1998. Bayesian PCA. In: Kearns, M., Solla, S., Cohn, D. (Eds.), Advances in Neural Information Processing Systems. MIT Press.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer, New York. Information Science and Statistics.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H.M., 2007. Forecasting the global burden of Alzheimer's disease. Alzheimer's Dement. 3, 186–191. https://doi.org/10.1016/j.jalz.2007.04.381.

Coley, N., Gardette, V., Cantet, C., Gillette-Guyonnet, S., Nourhashemi, F., Vellas, B., Andrieu, S., 2011. How should we deal with missing data in clinical trials involving Alzheimers disease patients? Curr. Alzheimer Res. 8, 421–433. https://doi.org/10.2174/156720511795745339.

Donohue, M.C., Jacqmin-Gadda, H., Le Goff, M., Thomas, R.G., Raman, R., Gamst, A.C., Beckett, L.A., Jack, C.R., Weiner, M.W., Dartigues, J.F., Aisen, P.S., 2014. Estimating long-term multivariate progression from short-term data. Alzheimer's Dement. 10 https://doi.org/10.1016/j.jalz.2013.10.003.

Fisher, C.K., Smith, A.M., Walsh, J.R., 2019. Coalition against major diseases, abbott, alliance for aging research, Alzheimer's Association, Alzheimer's Foundation of

America, Astrazeneca Pharmaceuticals LP, Bristol-Myers Squibb Company, Critical Path Institute, CHDI Foundation, Inc., Eli Lilly and Company, F. Hoffmann-La Roche Ltd, Forest Research Institute, Genentech, Inc., Glaxosmithkline, Johnson & Johnson, National Health Council, novartis Pharmaceuticals Corporation, Parkinson's Action Network, Parkinson's Disease Foundation, Pfizer, Inc., Sanofi-Aventis. Collaborating Organizations: Clinical Data Interchange Standards Consortium (CDISC), Ephibian, Metrum Institute. Machine learning for comprehensive forecasting of Alzheimer's Disease progression. Sci. Rep. 9, 13622. https://doi.org/10.1038/s41598-019-49656-2.

Fonteijn, H.M., Modat, M., Clarkson, M.J., Barnes, J., Lehmann, M., Hobbs, N.Z., Scahill, R.I., Tabrizi, S.J., Ourselin, S., Fox, N.C., Alexander, D.C., 2012. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. Neuroimage 60, 1880–1889. https://doi.org/10.1016/j.neuroimage.2012.01.062.

Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D., 2016. Instantiated mixed effects modeling of Alzheimer's disease markers. Neuroimage 142, 113–125. https://doi.org/10.1016/j.neuroimage.2016.06.049.

Hardy, S.E., Allore, H., Studenski, S.A., 2009. Missing data: a special challenge in aging research: MISSING DATA. J. Am. Geriatr. Soc. 57, 722–729. https://doi.org/10.1111/j.1532-5415.2008.02168.x.

Jack, C.R., Bennett, D.A., Blennow, K., Carrillo, M.C., Dunn, B., Haeberlein, S.B., Holtzman, D.M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J.L., Montine, T., Phelps, C., Rankin, K.P., Rowe, C.C., Scheltens, P., Siemers, E., Snyder, H.M., Sperling, R., Contributors, Elliott, C., Masliah, E., Ryan, L., Silverberg, N., 2018. NIA-AA Research framework: toward a biological definition of Alzheimer's disease. Alzheimer's Dement. 14, 535–562. https://doi.org/10.1016/j.jalz.2018.02.018.

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

Jedynak, B.M., Lang, A., Liu, B., Katz, E., Zhang, Y., Wyman, B.T., Raunig, D., Jedynak, C.P., Caffo, B., Prince, J.L., 2012. A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease neuroimaging initiative cohort. Neuroimage 63, 1478–1486. https://doi.org/10.1016/j.neuroimage.2012.07.059.

Koval, I., Bône, A., Louis, M., Lartigue, T., Bottani, S., Marcoux, A., Samper-González, J., Burgos, N., Charlier, B., Bertrand, A., Epelbaum, S., Colliot, O., Allassonnière, S., Durrleman, S., 2021. AD course map charts Alzheimer's disease progression. Sci. Rep. 11, 8020. https://doi.org/10.1038/s41598-021-87434-1.

Kühnel, L., Berger, A.K., Markussen, B., Raket, L.L., 2021. Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales. Stat. Med. 40, 3251–3266. https://doi.org/10.1002/sim.8932.

Kumar, S., Oh, I., Schindler, S., Lai, A.M., Payne, P.R.O., Gupta, A., 2021. Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review. JAMIA Open 4, ooab052. https://doi.org/10.1093/jamiaopen/ooab052.

Li, D., Iddi, S., Thompson, W.K., Donohue, M.C., Alzheimer Disease Neuroimaging Initiative, 2019. Bayesian latent time joint mixed effect models for multicohort longitudinal data. Stat. Methods Med. Res. 28, 835–845. https://doi.org/10.1177/0962280217737566.

López, M., Ramírez, J., Górriz, J., Álvarez, I., Salas-Gonzalez, D., Segovia, F., Chaves, R., 2009. SVM-based CAD system for early detection of the Alzheimer's disease using kernel PCA and LDA. Neurosci. Lett. 464, 233–238. https://doi.org/10.1016/j.neulet.2009.08.061.

Lorenzi, M., Filippone, M., Frisoni, G.B., Alexander, D.C., Ourselin, S., 2019. Probabilistic disease progression modeling to characterize diagnostic uncertainty: application to staging and prediction in alzheimer's disease. Neuroimage 190, 56–68. https://doi.org/10.1016/j.neuroimage.2017.08.059.

Mattsson, N., Carrillo, M.C., Dean, R.A., Devous, M.D., Nikolcheva, T., Pesini, P., Salter, H., Potter, W.Z., Sperling, R.S., Bateman, R.J., Bain, L.J., Liu, E., 2015. Revolutionizing Alzheimer's disease and clinical trials through biomarkers. Alzheimer's Dement. 1, 412–419. https://doi.org/10.1016/j.dadm.2015.09.001.

Mehdipour Ghazi, M., Nielsen, M., Pai, A., Modat, M., Jorge Cardoso, M., Ourselin, S., Sørensen, L., 2021. Robust parametric modeling of Alzheimer's disease progression. Neuroimage 225, 117460. https://doi.org/10.1016/j.neuroimage.2020.117460.

Nichols, E., Steinmetz, J.D., Vollset, S.E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T.T., Al Hamad, H., Alahdab, F., Alanezi, F.M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., Astell-Burt, T., Ayano, G., Ayuso-Mateos, J.L., Baig, A.A., Barnett, A., Barrow, A., Baune, B.T., Béjot, Y., Bezabhe, W.M.M., Bezabih, Y.M., Bhagavathula, A.S., Bhaskar, S., Bhattacharyya, K., Bijani, A., Biswas, A., Bolla, S.R., Boloor, A., Brayne, C., Brenner, H., Burkart, K., Burns, R.A., Cámera, L.A., Cao, C., Carvalho, F., Castro-de Araujo, L.F.S., Catalá-López, F., Cerin, E., Chavan, P.P., Cherbuin, N., Chu, D.T., Costa, V.M., Couto, R.A.S., Dadras, O., Dai, X., Dandona, L., Dandona, R., De la Cruz-Góngora, V., Dhamnetiya, D., Dias da Silva, D., Diaz, D., Douiri, A., Edvardsson, D., Ekholuenetale, M., El Sayed, I., El-Jaafary, S.I., Eskandari, K., Eskandarieh, S., Esmaeilnejad, S., Fares, J., Faro, A., Farooque, U., Feigin, V.L., Feng, X., Fereshtehnejad, S.M., Fernandes, E., Ferrara, P., Filip, I., Fillit, H., Fischer, F., Gaidhane, S., Galluzzo, L., Ghashghaee, A., Ghith, N., Gialluisi, A., Gilani, S.A., Glavan, I.R., Gnedovskaya, E.V., Golechha, M., Gupta, R., Gupta, V.B., Gupta, V.K., Haider, M.R., Hall, B.J., Hamidi, S., Hanif, A., Hankey, G.J., Haque, S., Hartono, R.K., Hasaballah, A.I., Hasan, M.T., Hassan, A., Hay, S.I., Hayat, K., Hegazy, M.I., Heidari, G., Heidari-Soureshjani, R., Herteliu, C., Househ, M., Hussain, R., Hwang, B.F., Iacoviello, L., Iavicoli, I., Ilesanmi, O.S., Ilic, I. M., Ilic, M.D., Irvani, S.S.N., Iso, H., Iwagami, M., Jabbarinejad, R., Jacob, L., Jain, V., Jayapal, S.K., Jayawardena, R., Jha, R.P., Jonas, J.B., Joseph, N., Kalani, R., Kandel, A., Kandel, H., Karch, A., Kasa, A.S., Kassie, G.M., Keshavarz, P., Khan, M.A., Khatib, M.N., Khoja, T.A.M., Khubchandani, J., Kim, M.S., Kim, Y.J., Kisa, A., Kisa, S., Kivimäki, M., Koroshetz, W.J., Koyanagi, A., Kumar, G.A., Kumar, M., Lak, H.M., Leonardi, M., Li, B., Lim, S.S., Liu, X., Liu, Y., Logroscino, G., Lorkowski, S., Lucchetti, G., Lutzky Saute, R., Magnani, F.G., Malik, A.A., Massano, J., Mehndiratta, M.M., Menezes, R.G., Meretoja, A., Mohajer, B., Mohamed Ibrahim, N., Mohammad, Y., Mohammed, A., Mokdad, A.H., Mondello, S., Moni, M. A.A., Moniruzzaman, M., Mossie, T.B., Nagel, G., Naveed, M., Nayak, V.C., Neupane Kandel, S., Nguyen, T.H., Oancea, B., Otstavnov, N., Otstavnov, S.S., Owolabi, M.O., Panda-Jonas, S., Pashazadeh Kan, F., Pasovic, M., Patel, U.K., Pathak, M., Peres, M.F. P., Perianayagam, A., Peterson, C.B., Phillips, M.R., Pinheiro, M., Piradov, M.A., Pond, C.D., Potashman, M.H., Pottoo, F.H., Prada, S.I., Radfar, A., Raggi, A., Rahim, F., Rahman, M., Ram, P., Ranasinghe, P., Rawaf, D.L., Rawaf, S., Rezaei, N., Rezapour, A., Robinson, S.R., Romoli, M., Roshandel, G., Sahathevan, R., Sahebkar, A., Sahraian, M.A., Sathian, B., Sattin, D., Sawhney, M., Saylan, M., Schiavolin, S., Seylani, A., Sha, F., Shaikh, M.A., Shaji, K., Shannawaz, M., Shetty, J. K., Shigematsu, M., Shin, J.I., Shiri, R., Silva, D.A.S., Silva, J.P., Silva, R., Singh, J.A., Skryabin, V.Y., Skryabina, A.A., Smith, A.E., Soshnikov, S., Spurlock, E.E., Stein, D. J., Sun, J., Tabarés-Seisdedos, R., Thakur, B., Timalsina, B., Tovani-Palone, M.R., Tran, B.X., Tsegaye, G.W., Valadan Tahbaz, S., Valdez, P.R., Venketasubramanian, N., Vlassov, V., Vu, G.T., Vu, L.G., Wang, Y.P., Wimo, A., Winkler, A.S., Yadav, L., Yahyazadeh Jabbari, S.H., Yamagishi, K., Yang, L., Yano, Y., Yonemoto, N., Yu, C., Yunusa, I., Zadey, S., Zastrozhin, M.S., Zastrozhina, A., Zhang, Z.J., Murray, C.J.L., Vos, T., 2022. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the global burden of disease study 2019. Lancet Public Health 7, e105–e125. https://doi.org/10.1016/S2468-2667(21)00249-8.

Porta, J., Verbeek, J., Kröse, B., 2005. Active appearance-based robot localization using stereo vision. Auton. Robots 18, 59–80. https://doi.org/10.1023/B:AURO.0000047287.00119.b6.

Raket, L.L., 2020. Statistical disease progression modeling in alzheimer disease. Front. Big Data 3, 24. https://doi.org/10.3389/fdata.2020.00024.

Rudovic, O.O., Utsumi, Y., Guerrero, R., Peterson, K., Rueckert, D., Picard, R.W., 2019. Meta-weighted Gaussian process experts for personalized forecasting of ad cognitive changes. In: Doshi-Velez, F., Fackler, J., Jung, K., Kale, D., Ranganath, R., Wallace, B., Wiens, J. (Eds.), Proceedings of the 4th Machine Learning for Healthcare Conference. PMLR, pp. 181–196.https://proceedings.mlr.press/v106/rudovic19a.html.

Scheltens, P., De Strooper, B., Kivipelto, M., Holstege, H., Chételat, G., Teunissen, C.E., Cummings, J., van der Flier, W.M., 2021. Alzheimer's disease. The Lancet 397, 1577–1590. https://doi.org/10.1016/S0140-6736(20)32205-4.

Schindler, S.E., Li, Y., Buckles, V.D., Gordon, B.A., Benzinger, T.L., Wang, G., Coble, D., Klunk, W.E., Fagan, A.M., Holtzman, D.M., Bateman, R.J., Morris, J.C., Xiong, C., 2021. Predicting symptom onset in sporadic alzheimer disease with amyloid PET. Neurology 97, e1823–e1834. https://doi.org/10.1212/WNL.0000000000012775.

Schiratti, J.B., Allassonnière, S., Routier, A., Initiative, t.A.D.N., Colliot, O., Durrleman, S., 2015. A mixed-effects model with time reparametrization for longitudinal univariate manifold-valued data. In: Ourselin, S., Alexander, D.C., Westin, C.F., Cardoso, M.J. (Eds.), Information Processing in Medical Imaging, Vol. 9123. Springer International Publishing, Cham, pp. 564–575. https://doi.org/10.1007/978-3-319-19992-4_44.

Shan, G., Bernick, C., Caldwell, J.Z.K., Ritter, A., 2021. Machine learning methods to predict amyloid positivity using domain scores from cognitive tests. Sci. Rep. 11, 4822. https://doi.org/10.1038/s41598-021-83911-9.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Figurski, M., Coart, E., Blennow, K., Soares, H., Simon, A.J., Lewczuk, P., Dean, R.A., Siemers, E., Potter, W., Lee, V.M.Y., Trojanowski, J.Q., Initiative, t.A.D.N., 2011. Qualification of the analytical and clinical performance of CSF biomarker analyses in ADNI. Acta Neuropathol. 121, 597–609. https://doi.org/10.1007/s00401-011-0808-0.

Smith, S.M., Vidaurre, D., Alfaro-Almagro, F., Nichols, T.E., Miller, K.L., 2019. Estimation of brain age delta from brain imaging. Neuroimage 200, 528–539. https://doi.org/10.1016/j.neuroimage.2019.06.017.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Kaye, J., Montine, T.J., Park, D.C., Reiman, E.M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M.C., Thies, B., Morrison-Bogorad, M., Wagster, M.V., Phelps, C.H., 2011. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. Alzheimer's Dement. 7, 280–292. https://doi.org/10.1016/j.jalz.2011.03.003.

Tabarestani, S., Aghili, M., Shojaie, M., Freytes, C., Adjouadi, M., 2018. Profile-specific regression model for progression prediction of Alzheimer's disease using longitudinal data. 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, Orlando, FL, pp. 1353–1357. https://doi.org/10.1109/ICMLA.2018.00220.

Tami, M., Clausel, M., Devijver, E., Gaussier, E., Aubert, J.M., Chebre, M., 2018. Decision tree for uncertainty measures. JDS 2018 : 50èmes Journées de Statistique. Paris-Saclay, France.

Tipping, M.E., Bishop, C.M., 1999. Probabilistic principal component analysis. J. R. Stat. Soc. Ser. B (Statistical Methodology) 61, 611–622. https://doi.org/10.1111/1467-9868.00196.

Venkatraghavan, V., Bron, E.E., Niessen, W.J., Klein, S., 2019. Disease progression timeline estimation for Alzheimer's disease using discriminative event based modeling. Neuroimage 186, 518–532. https://doi.org/10.1016/j.neuroimage.2018.11.024.

Voevodskaya, O., Simmons, A., Nordenskjöld, R., Kullberg, J., Ahlström, H., Lind, L., Wahlund, L.O., Larsson, E.M., Westman, E., Initiative, A.D.N., 2014. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. Front. Aging Neurosci. 6 https://doi.org/10.3389/fnagi.2014.00264.

Wang, X., Qi, J., Yang, Y., Yang, P., 2019. A survey of disease progression modeling techniques for Alzheimer's diseases. 2019 IEEE 17th International Conference on Industrial Informatics (INDIN). IEEE, Helsinki, Finland, pp. 1237–1242. https://doi.org/10.1109/INDIN41052.2019.8972091.

Yagi, T., Kanekiyo, M., Ito, J., Ihara, R., Suzuki, K., Iwata, A., Iwatsubo, T., Aoshima, K., Initiative, A.D.N., Initiative, J.A.D.N., 2019. Identification of prognostic factors to predict cognitive decline of patients with early Alzheimer's disease in the Japanese Alzheimer's disease neuroimaging initiative study. Alzheimer's & Dement. 5, 364–373. https://doi.org/10.1016/j.trci.2019.06.004.

Yang, E., Farnum, M., Lobanov, V., Schultz, T., Raghavan, N., Samtani, M.N., Novak, G., Narayan, V., DiBernardo, A., Initiative, t.A.D.N., 2011. Quantifying the pathophysiological timeline of Alzheimer's disease. J. Alzheimers Dis. 26, 745–753. https://doi.org/10.3233/JAD-2011-110551.

Young, A.L., Marinescu, R.V., Oxtoby, N.P., Bocchetta, M., Yong, K., Firth, N.C., Cash, D.M., Thomas, D.L., Dick, K.M., Cardoso, J., van Swieten, J., Borroni, B., Galimberti, D., Masellis, M., Tartaglia, M.C., Rowe, J.B., Graff, C., Tagliavini, F., Frisoni, G.B., Laforce, R., Finger, E., de Mendonça, A., Sorbi, S., Warren, J.D., Crutch, S., Fox, N.C., Ourselin, S., Schott, J.M., Rohrer, J.D., Alexander, D.C., (GENFI), T.G.F.I., (ADNI), T.A.D.N.I., 2018. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. Nat. Commun. 9, 4273. https://doi.org/10.1038/s41467-018-05892-0.

Young, A.L., Oxtoby, N.P., Daga, P., Cash, D.M., Fox, N.C., Ourselin, S., Schott, J.M., Alexander, D.C., 2014. A data-driven model of biomarker changes in sporadic Alzheimer's disease. Brain 137, 2564–2577. https://doi.org/10.1093/brain/awu176.