scientific reports

OPEN

AD Course Map charts Alzheimer's disease progression

Igor Koval[1,2,5,7], Alexandre Bône[1,2,7], Maxime Louis[1,2], Thomas Lartigue[1,2,5], Simona Bottani[1,2], Arnaud Marcoux[1,2], Jorge Samper-González[1,2], Ninon Burgos[1,2], Benjamin Charlier[1,2,6], Anne Bertrand[1,2,3,8], Stéphane Epelbaum[1,2,3], Olivier Colliot[1,2,3], Stéphanie Allassonnière[4,5] & Stanley Durrleman[1,2]✉

Alzheimer's disease (AD) is characterized by the progressive alterations seen in brain images which give rise to the onset of various sets of symptoms. The variability in the dynamics of changes in both brain images and cognitive impairments remains poorly understood. This paper introduces AD Course Map a spatiotemporal atlas of Alzheimer's disease progression. It summarizes the variability in the progression of a series of neuropsychological assessments, the propagation of hypometabolism and cortical thinning across brain regions and the deformation of the shape of the hippocampus. The analysis of these variations highlights strong genetic determinants for the progression, like possible compensatory mechanisms at play during disease progression. AD Course Map also predicts the patient's cognitive decline with a better accuracy than the 56 methods benchmarked in the open challenge TADPOLE. Finally, AD Course Map is used to simulate cohorts of virtual patients developing Alzheimer's disease. AD Course Map offers therefore new tools for exploring the progression of AD and personalizing patients care.

Alzheimer's disease (AD) is an extraordinarily complex disease. Its effects on the brain are visible on multiple radiological examinations. Their consequences on the cognition and behavior are evaluated by a series neuro-psychological assessments. Adding to this complexity, the disease develops during decades of life, the progression is not linear, and the timing among radiological and clinical signs greatly varies across individuals.

For all these reasons, it is difficult for the clinicians to predict how the disease will progress in each patient and to forecast how the progression of one patient will differ from another one. This difficulty hampers the identification of the optimal moment to test a therapeutic intervention and a fine-grained description of how the drug modifies the different aspects of disease progression.

In this paper, we propose a system to position any patient at any time-point in a map of disease progression, called AD Course Map. AD Course Map summarizes the range of likely trajectories followed by patients developing AD. The positioning includes the identification a precise disease stage and a patient specific trajectory showing how the patient's data will change in the coming years. The positioning of several subjects opens up the possibility to explore the variability of disease manifestation via a small number of interpretable parameters.

We construct AD Course Map using a statistical learning algorithm and a longitudinal multimodal observational data set. The statistical analysis of longitudinal data is challenging because standard methods do not account for age at onset and pace of progression as possible co-founding factors[1]. In fact, data from different subjects may differ at least two reasons. First, data vary because the disease progresses with different dynamics. Even two patients showing a very typical form of the disease with the same sequence of radiological and clinical signs may develop it at a different age and at a different pace. Even if they follow the same trajectory, the data of these two patients taken at the same age still differ: patients are not at the same disease stage. Second, data vary because the presentation of the disease varies according to patients. One patient may show a predominantly amnesic form of the disease, while another one show primarily attention deficits. Therefore, even if both patients are observed at the same disease stage, the assessments of their memory and attention capacity still differ. The statistical analysis should disentangle such phenotypic differences from differences in the dynamics of progression.

[1]Institut du Cerveau et de la Moelle épinière (ICM) & Inserm, U 1127 & CNRS, UMR 7225, Sorbonne Université, 75013 Paris, France. [2]Inria, Aramis project-team, Paris, France. [3]AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France. [4]Centre de Recherche des Cordeliers, Université Paris Descartes, Paris, France. [5]Centre de Mathématiques Appliquées, Ecole Polytechnique, Palaiseau, France. [6]Laboratoire Alexandre Grotendieck, Université de Montpellier, Montpellier, France. [7]These authors contributed equally: Igor Koval and Alexandre Bône. [8]Anne Bertrand is deceased. ✉email: stanley.durrleman@inria.fr
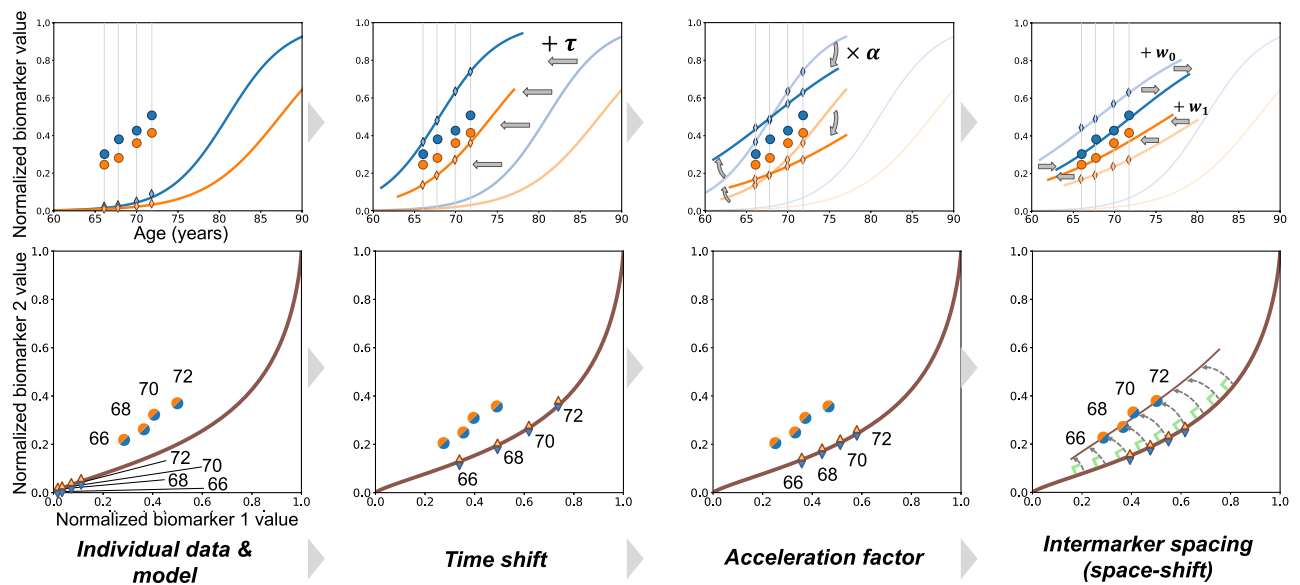
Figure 1. Disease course mapping for two biomarker data. Top left panel: the model as plain curves and repeated data of one subject. x-axis is age in years and y-axis is the normalized values of the biomarkers. Top row, three left panels: the three operations used to mapping the model to the individual data: the time-shift translates the curves, the acceleration factor scales the abscissa, the space-shifts change the time interval between both curves. Bottom row shows another representation of the *same* data as parametric curves. Panels plot the values of one biomarker versus the other one, time being the parameter. Data fall within a unit square, which is a particular case of a Riemannian manifold. The model is a geodesic curve and the transformed curve is a change in the parametric representation of the geodesic followed by exp-parallelisation, a generalisation of translation on manifolds. This construction for two biomarker data extends therefore to any kind of data on a Riemannian manifold.

Describing the differences in the dynamics of progression has been tackled in previous disease modeling techniques using functions that map the actual age of the patient to a stage of progression. These functions have been introduced as time-warps[2,3] and have been included in latent time models[4,5]. Several statistical models assume that all patients follow the same trajectory but with a different dynamics[6–8]. Data take always the same values but not at the same age. These approaches ignore possible differences in disease presentation. Conversely, event-based models[9,10] explores how the temporal sequence of events varies, the event being defined as the moment when a biomarker becomes abnormal. Nevertheless, this approach does not link the temporal sequence with the age of the patient.

Combining dynamic and phenotypic variations in a statistical model is difficult because their effects may be confounded. Disease course mapping is a statistical technique which allows the decomposition of these two types of variations seen in a longitudinal data set. It builds on abstract geometric principles so that it is not specific to particular data[11,12]. It combines the concept of time-warp with translations of curves.

This technique extends the concept of atlas in neuroimaging to locate anatomical structures in brain images independently of their variations in size and overall shape of the brain. Each anatomical structure has its own set of (x,y,z) coordinate in a reference image of the brain. The reference image is transformed so that it superimposes with the image of the subject. This transformation puts into correspondence each (x,y,z) coordinate of the atlas to a point in the image of the subject. It locates therefore the anatomical structures in the subject's brain. If one repeats this procedure for a set of representative subjects, the transformations applied to the atlas image describe the variability in size and shape of the anatomical structures within the population.

Disease course mapping extends this concept for the progression of a series of biomarkers. The atlas takes the form a series of progression curves for each biomarker where the time axis represents a disease stage, which we call here an Alzheimer Age (AA). Subject's data take the form the biomarkers measurements at several time-points. The transformation of model to the subject's data consist in computing (1) a time-shift, which translates the model along the time axis to accommodate for changes in age at disease onset, (2) an acceleration factor, which scales the time interval to accommodate for differences in speed of progression, and (3) intermarker spacings, which translate each biomarker differently to accommodate for differences in the timing and ordering among biomarkers. The first two parameters define a time-warp function that maps the actual age of the subject to his AA. This function therefore changes the dynamics of progression but not the trajectory. In turn, inter-marker spacings shift the patient trajectory to account for phenotypic differences across subjects, as illustrated in Figure 1.

Disease course mapping is not specific to biomarker data. It further extends to the case where data are images or meshes of anatomical structures. The reference scenario is replaced by a scenario of temporal changes of image intensities or a scenario of continuous deformation of a shape respectively. Time-shifts and acceleration factor plays the same role of adjusting for the dynamics of change. They map the actual age of the subject at each visit to its Alzheimer Age that is a time-point in the reference scenario. Intermarker spacing adjust for the spatiotemporal
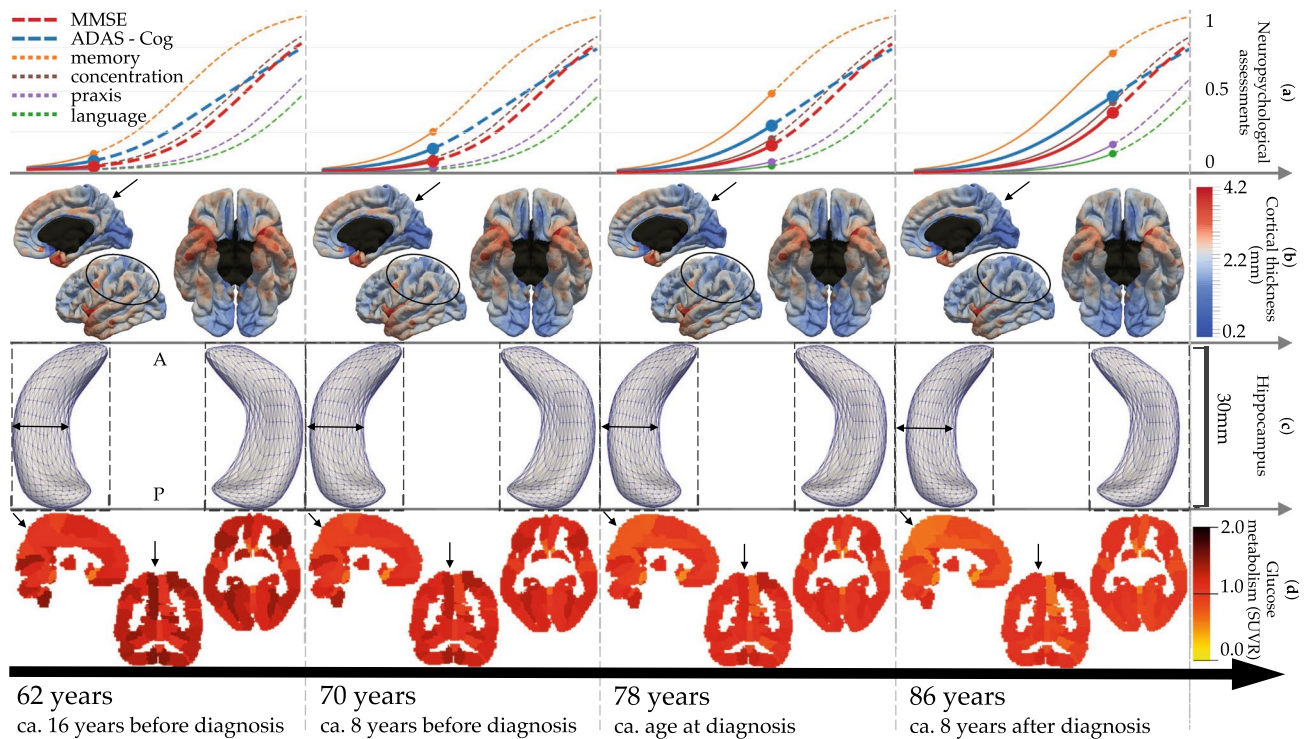
Figure 2. Normative models of Alzheimer's disease progression shown at 4 Alzheimer Age with estimated time until/from diagnosis. Bottom to top rows show alteration of brain glucose metabolism, hippocampus atrophy, cortical thinning and onset of cognitive decline. Black arrows and ellipses indicate some areas of great changes. Images were obtained using the freely available software FSLeyes v.0.22.6 (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLeyes) and Paraview v.5.2.0 (www.paraview.org).

pattern of changes across brain regions or for the shape of the structure respectively. In this setting intermarker spacing takes the more general name of space-shift.

We use here a longitudinal data sets of patients who were diagnosed with Alzheimer's disease at some point of the study to reconstruct the natural history of the disease. We construct an AD Course Map showing the reduction in glucose consumption across brain regions, the thinning of the cerebral cortex, the deformation of the hippocampus due to atrophy, and the progressive onset of cognitive deficit from the preclinical to the clinical phase of the disease.

We analyse the distribution of the parameters that adjust the reference scenario of disease progression to each subject's data. This statistical analysis aims to highlight how genetic, biological or environmental factors change disease progression profiles. Once the map is built, we position new individuals into the map and then read how data will change in the coming years. Eventually, we build a virtual cohort of patients developing AD by sampling random trajectories in the map.

Results

AD Course Map: a multimodal atlas of disease progression.
We estimated AD Course Map using multimodal longitudinal data from subjects covering the pre-clinical to the clinical stage of the disease (see Methods). Figure 2 shows at four representative Alzheimer ages (AA) the normative scenario of cognitive decline (first row), spatiotemporal propagation of cortical thinning across cortical regions (second row), progressive deformation of the hippocampus in both hemispheres due to atrophy (third row), and spatiotemporal progression hypometabolism across brain regions (fourth row). AD Course Map may be visualised at a fine temporal resolution in the form of an interactive visualisation at the website: www.digital-brain.org.

We found that the AA at the time of diagnosis is of 78 (±5.6) years. The model represented in Fig. 2 encompasses therefore 16 years before diagnosis and 8 years after. It has been shown that Alzheimer's disease diagnosis occurs when the ADAS-Cog is comprised between 18.6 and 28.9 (i.e. between 0.21 and 0.34 on the normalised scale)[13], which is reached in the AD Course Map for an AA between 74 and 80 years. Similarly, the diagnosis usually occurs for a MMSE score comprised between 27 and 23 (i.e. 0.1 and 0.23 on the normalised scale)[14], which occurs in the AD Course Map for an AA between 74 and 81 years. These age ranges are compatible with our estimation of the diagnosis occurring at an AA of 78 (±5.6) years.

AD Course Map shows a typical sequence of cognitive impairments starting with memory, followed by concentration 9.6 (±1.54) years after, praxis 9.8 (±1.73) years after, and finally language 3.3 (±2.65) years after. The greatest alterations of glucose hypometabolism are found in the precuneus[15–17], prefrontal areas[18] and the parahippocampal region[19]. Cortical atrophy also occurs in typical regions such as enthorinal cortex, hippocampal gyrus, temporal pole and fusiform gyrus[20,21], cortical association areas[22,23] and precuneus[24]. As expected, very

(a) FDG-PET images

(b) Cortical thickness maps

(c) Left hippocampus mesh   **(d)** Right hippocampus mesh   **(e)** Neuropsychological assessments
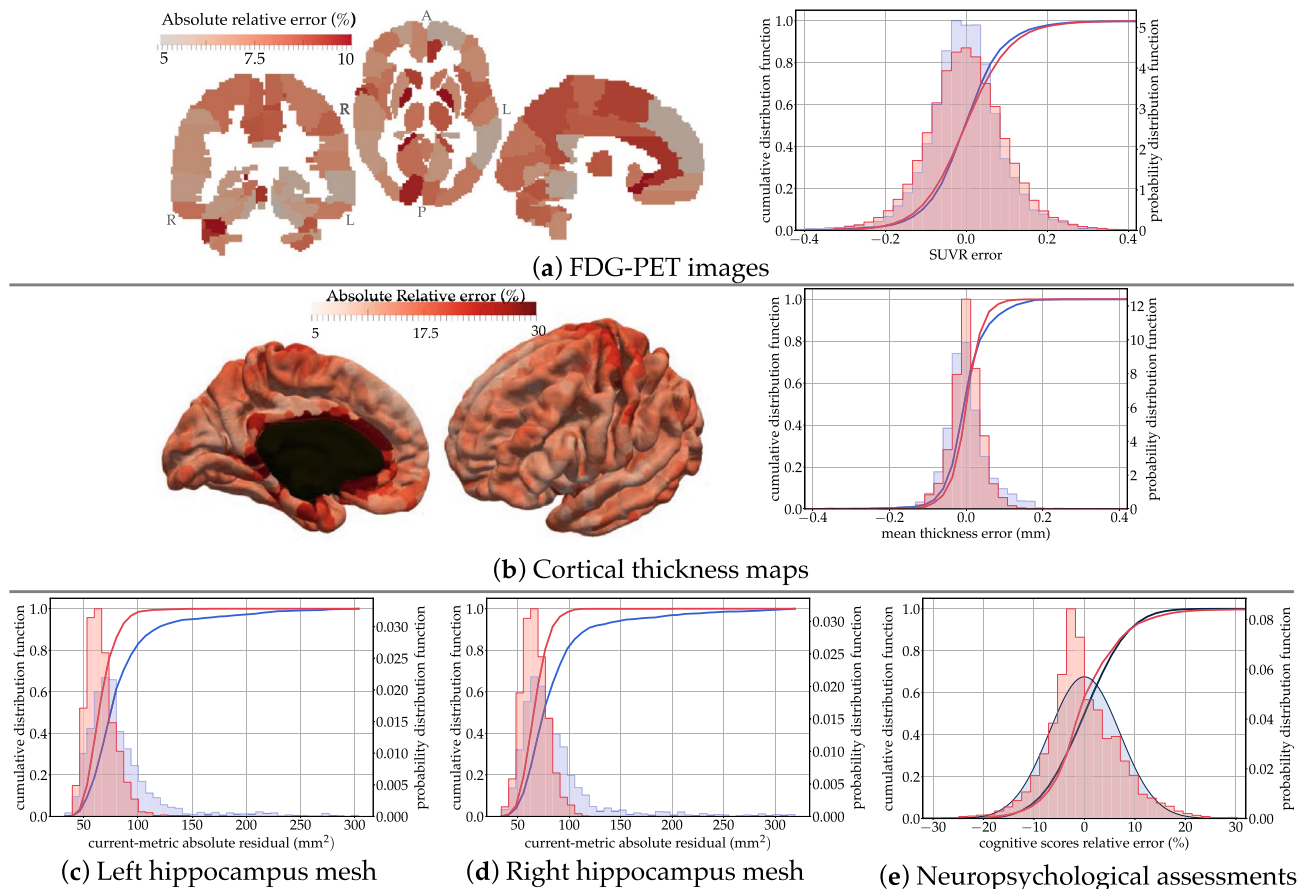
Figure 3. Distributions of reconstruction errors. The empirical distribution of errors (red) is superimposed with the estimated distribution of test / re-test differences (in blue). The absolute relative error is shown in every brain region for FDG-PET images and cortical thickness maps. Mean and standard errors are given in Supplementary Table S1. Images were obtained using the freely available software FSLeyes v.0.22.6 (https://fsl. fmrib.ox.ac.uk/fsl/fslwiki/FSLeyes) and Paraview v.5.2.0 (www.paraview.org).

little atrophy is shown to occur in the occipital lobe and the cingulate gyrus. These structural and functional alterations are in line with previous findings. More surprisingly, the model shows atrophy in the precentral gyrus and the paracentral lobule. Whether these regions are affected by cortical thinning due to Alzheimer's disease is still a debated question[25]. We found that the level of noise in these regions is one of the largest, a fact that may explain inconsistencies across studies.

AD Course Map fits data with an error of the order of the measurement uncertainty. AD Course Map estimates not only a typical progression of clinical and imaging data, but also a range of possible trajectories. These trajectories derive from the typical scenario by adjusting the age at onset, pace of progression, appearance of the images, the shape of meshes, and the temporal spacing among clinical assessments.

After model calibration, each subject is positioned on one of these possible trajectories. The positioning amounts to the estimation of subject-specific time-shift, acceleration factor and a series of space-shift for each modality. The space-shifts determine the subject-specific trajectory. The time-shift and acceleration factor allow the computation of the Alzheimer Age of the subject, and therefore locate one point on the trajectory at each visit. We measure now to which extend the data at these points on the trajectory are similar to the observed data. We call the difference the *reconstruction error*.

Figure 3 shows the distribution of the reconstruction errors for all visits from all subjects. We notice that the reconstruction errors in brain regions are not evenly distributed. For PET data, the largest errors are found mostly in smaller regions. For cortical thickness, larger errors are found at the boundary of the mesh with the corpus callosum, mostly due to interpolation errors. These errors are much smaller than the best possible image resolution of 1 mm isotropic, thus making these reconstructions at sub-voxel precision. Reconstruction error for the hippocampus is reported in terms of the current distance[26], which is measure of difference in shape between two meshes. It is robust to mesh imperfection, spikes and holes. Errors in terms of volume are reported in Supplementary Figure S1.

We further evaluate how good is the goodness-of-fit by estimating the uncertainty of the measurements due to acquisition and processing. Imaging data are subject to variations in the acquisition protocol and experimental conditions such as slight motion of the subject in the scanner. Clinical assessments are subject to inter-rater

variability and subject's condition. Research protocols tend to reduce these variations as much as possible. Nevertheless, data are known only within an error margin which is difficult to evaluate. For MRI data, we estimate this uncertainty by measuring differences between test and re-test MRI sessions, two images acquired on the same subject one right after the other. For PET data, there is no such test / re-test data. As a proxy, we measured the difference between the baseline and follow-up image after 18 months for amyloid negative cognitively normal subjects. It is an over-estimation of the measurement uncertainty as this proxy neglects the effect of normal aging on the change in FDG uptake. Nevertheless, studies have not shown statistically significant longitudinal changes in FDG uptake over such a short period of time for normal ageing individuals[27,28]. Another study reports significant longitudinal changes over an average period of 7.8 years in the anterior and posterior cingulate cortex with annual rates of change of 0.58% and 0.57% respectively[29]. For the neuropsychological assessments, we perform a review of the literature assessing their reproducibility (see Methods). From these estimates, we deduce an empirical distribution of the noise in the data.

Fig. 3 shows the superimposition of the empirical distribution of reconstruction errors with the empirical distribution of the noise for all data types. Overall, the two distributions largely overlap, and the standard error is of the same order than the measurement noise (see Supplementary Table S1). This result shows that our model explains the variance seen in the data up to the noise level. Therefore, for the metric on the data that we used, these results could not be improved without over-fitting.

AD Course Map fits new subjects' data equally well. Once the AD Course Map has been calibrated on a representative data set, we aim to use it to analyse the data of new subjects. We estimate here the accuracy of AD Course Map to fit new subject's data.

We use a five-fold cross validation procedure. The model is calibrated using 80% of the subjects, and the remaining 20% subjects are then positioned on the map. In the positioning phase, we use the typical scenario of progression and the covariance among subject specific parameters, which were estimated during the calibration phase. For each of the unseen subjects, we only estimate one time-shift, one acceleration factor and a series of space-shifts that minimize the discrepancy between the predicted data and the true subject's data. We call this discrepancy a *generalisation error*.

Distributions of the generalisation errors are essentially identical with these the reconstruction errors (see Supplementary Fig. S2). Only hippocampus shows a slightly higher generalisation error but still below the noise level estimated with test / re-test data. The goodness-of-fit is therefore equally good whether the subject has been included in the training set or in the test set.

AD Course Map calibration is robust against resampling. After the cross-validation procedure, each subject has been used exactly once as a test subject. We consider the subject-specific parameters that were estimated at that time. We compare them with the parameter values that were estimated when calibrating the model on the whole data set, and thus when the considered subject was used in the training set. The difference between the two estimates is small with r^2 comprised between 0.93 and 0.99 (see Supplementary Fig. S3).

Eventually, we use the cross-validation procedure to assess the robustness of AD Course Map against resampling. After the procedure, we have estimated five different course maps. Each course map is determined by the parameters of the typical model of progression and the variance of the subject-specific parameters. We compare these parameters estimated using 80% of the subjects with the ones resulting of the calibration of the model on the whole data set. These parameters show very little variations, thus showing the robustness of the estimation algorithm against resampling in the training set (see Supplementary Table S2).

AD Course Map summarizes the variety of possible disease progression profiles.
AD Course Map positions the progression of each subject with respect a reference. The time-shift and acceleration factor explain differences in age at onset and pace of progression. The set of space-shifts explain how different the subject' trajectory is in terms of image appearance, shape or timing among clinical assessments. Altogether these subject-specific parameters form the coordinates of the subject in a spatiotemporal coordinate system.

The set of these parameters describes the heterogeneity in disease progression profiles in the studied population. This analysis allows us to disentangle the variations in the dynamics of progression (how early or late is the disease onset of the subject? how fast or slow his progression?) from the variation in the disease presentation (does the subject's brain exhibit a particular shape, does it show a different ordering in the onset of the symptoms or in the functional alterations across brain regions?)

AD Course Map highlights possible compensatory mechanisms. For each modality, we perform a multivariate linear regression between each individual parameters and genetic, biological and environmental factors: sex, APOE-ε4 genotype, presence of amyloidosis, marital status and education level. We identify statistically significant associations using a two tailed t-test at 5% significance level corrected for multiple comparisons with the false discovery rate method (see Methods). Note that in this section, we discard subjects without assessments of amyloidosis (see Supplementary Table S3 for corresponding number of samples).

Our results in Table 1 show the predominant role of genetic factors to explain the heterogeneity in disease progression. In particular, disease progression presents a strong sexual dimorphism for hippocampus atrophy and cognitive decline. In women, the hippocampus atrophy occurs on average 33.6 months earlier and 1.27 times faster than in men. This atrophy translates into an earlier and accelerated cognitive decline which occurs 36.8 months earlier and 1.46 times faster in women than in men.

By contrast, APOE-ε4 carriers also exhibit earlier and more pronounced alterations of their hippocampus (45 months ealier and 1.17 times faster in carriers compared to non-carriers), but this effect is, to some extend,

			Hypometabolism (FDG-PET)	Hippocampus atrophy (MRI): Left hemisphere	Right hemisphere	Cortical thinning (MRI)	cognitive decline (ADAS+MMSE)
Genetic	sex female vs. male	Accel. factor		×1.27 CI = [1.11, 1.45] p = 2.26e−3**	×1.26 CI = [1.08, 1.45] p = 6.15 − 3**		×1.46 CI = [1.10, 1.92] p = 8.42e−3**
		Time-shift		−33.6 CI = [−55.8, −11.6] p = 3.71e−3**	−29.0 CI = [−53.0, −4.91] p = 2.31e−2*		−36.8 CI = [−62.0, −11.6] p = 4.48e−3**
		Space-shift		±0.55 CI = [0.28, 0.82] p = 4.00e−4***	±0.60 CI = [0.34, 0.86] p = 3.89e−5****	±0.48 CI = [0.22, 0.75] p = 2.24e−3**	
	APOE-ε4 carrier vs. non-carrier	Accel. factor		×1.17 CI = [1.02, 1.33] p = 2.77e−2*		×1.42 CI = [1.12, 1.82] p = 2.17e−2*	×1.25 CI = [1.03, 1.51] 2.17e−2*
		Time-shift		−45.0 CI = [−66.9, −23.2] p = 1.57e−4***	−36.8 CI = [−60.5, −13.0] p = 4.27e−3**		
		Space-shift					
Biological	amyloid positive vs. negative	Accel. factor		×1.18 CI = [1.06, 1.32] p = 8.20e−3**	×1.23 CI = [1.09, 1.39] p = 4.03e−3**		−21.9 CI = [−41.2, −2.5] p = 2.70e−2*
		Time-shift					
		Space-shift				±0.28 CI = [0.05, 0.50] p = 2.24e−3***	
Environmental	marital married vs. non-married	Accel. factor		×1.25 CI = [1.07, 1.48] p = 1.08e−2*			
		Time-shift		−59.5 CI = [−86.6, −32.5] p = 1.06e−4***	−52.7 CI = [−82.2, −23.2] p = 1.28e−3**		−32.6 CI = [1.8, 63.3] p = 3.78e−2*
		Space-shift					
	education nb. of years of education	Accel. factor					
		Time-shift		−6.04 CI = [−9.67, −2.42] p = 1.95e−3**	−7.60 CI = [−11.55, −3.64] p = 9.53e−4***		
		Space-shift					

Table 1. Significant associations of individual parameters with genetic, biological and environmental factors: effect sizes, confidence intervals at 95%, and adjusted p-values. Only adjusted p-values below 5% significance level are shown. Time-shifts are in months, other quantities have no units. Directions of space-shift are not signed. The figures on the top of the column "hippocampal atrophy" reads: "atrophy of the left hippocampus progresses 1.27 times faster in women than in men, starts 33.6 months earlier, and the hippocampus shape is significantly different between men and women regardless of the disease stage".
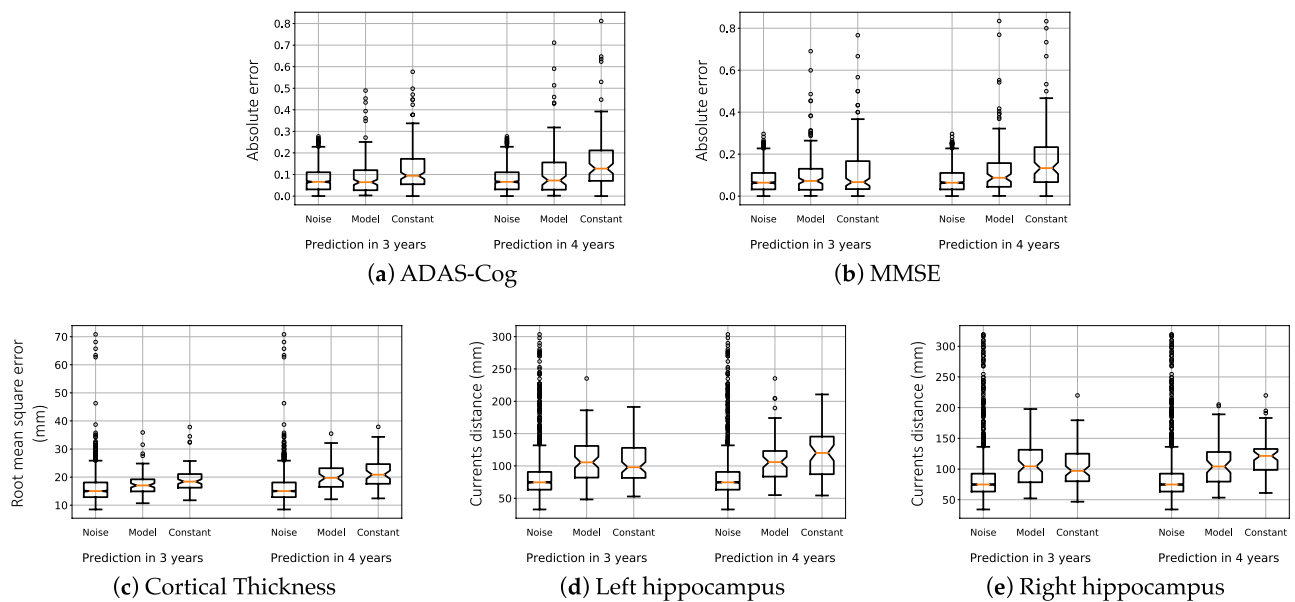
Figure 4. Prediction errors at the individual level 3 and 4 years ahead of time. Box-plots show medians in orange, quartiles, and 95% confidence intervals for three image data and the ADAS-Cog. Distributions of prediction errors are compared with that of the noise and the errors of the constant prediction.

alleviated in the onset of cognitive decline which does not occur earlier in carriers than in non-carriers. It is as if brain plasticity compensates for the advance of almost 3 years in hippocampal atrophy, but that once the compensation is made, cognitive decline still manifests itself at a faster rate than in subjects without the mutation.

The absence of associations between cofactors and profiles of hypometabolism may be explained by the fact that focal effects on specific brain areas may be diluted in non-specific regions of interest[30]. Except in four occasions, we found associations with parameters that modulate the dynamics of disease progression, not its presentation. This fact suggests that previous findings showing associations between genetic variants and hypometabolism might be due to the comparison of subjects at different ages or disease stages[30,31]. Age is a poor proxy of disease progression. Time from/to diagnosis alone does not allow to correct for variations in pace of progression. By contrast, we propose here a method to distinguish dynamic from phenotypic variations.

AD Course Map exhibits complex interplay between structural and functional alterations and cognitive decline. We construct a graph of a conditional correlations between all variables of all modalities. The statistically significant conditional correlations are represented in Figure 6 where three variables per modality are shown: pace of disease progression (e.g. acceleration factors), delay with respect to onset (e.g. time-shift), and pattern (e.g. space-shifts represented here as a single variable for the sake of simplicity).

Interestingly, the vast majority of significant conditional correlations are found among variables of the same type across modalities, and not among different variables within the same modality. It means that the three aspects of disease progression: pace of progression, age at onset and disease presentation are mostly independent of each other.

Cognitive decline is the consequence of the accumulation of lesions in the brain. Bearing this in mind, this graph confirms that the order and timing of the decline of the different cognitive functions is determined by the order and timing of the brain regions affected by neurodegeneration[32]. It further shows that the age at which these cognitive changes occur is associated partly to the age at which hippocampus atrophy occur, and not to the age at which hypometabolism and cortical atrophy starts. These weak to no association may be interpreted again as the consequence of possible compensatory mechanisms for a part of the patients.

We do not find association between pace of progression and age at onset, except for cortical atrophy. Studies reported that some early form of the disease are associated with more rapid progression[33,34], while others do not find any association[35,36]. Such association studies often use an arbitrary threshold on the diagnosis age to divide early from late onset AD patients. Within each group, a linear rate of change is computed, thus assuming that cognition declines in a linear fashion at all stages. By contrast, our method accounts for a non-linear dynamic of change. It positions patients along a continuous disease progression timeline without the need to cluster patients into arbitrary categories.

AD Course Map predicts the future progression of new subjects. We evaluate the ability of AD Course Map to predict the disease progression in the distant future for the subjects at risk of developing AD. For this purpose, we select subjects and visits in the ADNI database based on criteria that can be assessed from present and past visits only, without the need to know the whole disease history of the patients as previously.

We select all the visits of all the subjects in ADNI, which met the following conditions:

- the subject is labeled as Mild Cognitive Impairment at this visit,
- the MMSE of the subject is smaller or equal to 27 at this visit, as a confirmation of the MCI stage,
- the subject is amyloid positive at this visit,
- the sequence of diagnosis labels in the past visits is monotonic, meaning we exclude subjects showing reversion to cognitively normal, or having AD label in the past.

These criteria target patients with MCI due to AD according to NIA-AA criteria[37], a population of interest for clinical trials[38].

For each subject, we select the visits for which there is at least one follow-up of the same subject 3 or 4 years later in time. We hide the follow-up visits after the considered visit. We position the subject onto the AD Course Map using subject's data until the considered visit only. We read the subject's data at the next hidden visits on the subject's trajectory. We compare the predicted data at the time-point of the follow-up visits with the data that were hidden. We call this difference a "prediction error" (see Methods).

If the selected subject is not part of the previous training data set, we use the AD Course Map estimated in the previous section. If the selected subject is part of it, we use the AD Course Map that was estimated during the cross-validation procedure, using the fold in which the considered subject was a test subject.

We predict the MMSE and the ADAS-Cog for 136 subjects for the prediction at 3 years, and 80 subjects for the prediction at 4 years. We predict MRI data (cortical thickness maps and hippocampus shape) for 72 subjects at 3 years, and 63 subjects at 4 years. The prediction of the neuropsychological assessments was done using patient data at only one time-point in 36.5% of the cases at 3 years, and 39.1% at 4 years. Prediction of the MRI data used only one visit in 33.3% of the cases at 3 years and 58.9% at 4 years.

We report in Fig. 4 the error of prediction for the neuro-psychological assessments, the map of the cortical thickness and the shape of the hippocampus of both hemispheres. We compare the prediction errors with the distribution of the noise in the measurements and with the "constant" prediction, which predicts the data will not change in the next 3 or 4 years.

The mean absolute error (MAE) of the prediction of the MMSE is of 3.2 points at 3 years and 4 points at 4 years. These errors are of the same order as the accuracy of the test, which is of about 10%, so 3 points on a scale of 30. For ADAS-Cog, the MAE of 7.6 points at 3 years and 10.1 points at 4 years, where the 10% error range is of 8.5 points.

The error of the constant prediction, though increasing with the time-to-prediction, remains small in comparison with the measurement uncertainty up to 4 years in time. The difference in the median is of 0.001 at 3 years and 0.02 at 4 years for the MMSE normalized on a zero to one scale ($p < 0.05$ with a Man-Whitney U test), thus meaning an error smaller than 1 point of MMSE. It has already been shown that neuropsychological assessments, such as the MMSE, lack robustness to track longitudinal changes over short periods of time[39]. These results advocate for longer term predictions, and possibly more sensitive imaging markers[40].

For MRI derived data, the noise distribution presents a very heavy tail that is due to the large heterogeneity of the image quality and its consequence in data processing, at least using the standard acquisition protocol and processing pipelines of ADNI. The constant prediction does not show such a heavy tail, as images of best quality 3 or 4 years apart show less variability than the test and re-test image acquired the same day. Our model shows steady performance at 3 and 4 years, whereas the constant prediction worsens as time-to-prediction increases.

Comparing such predictions with alternative methods is difficult, as they are often evaluated on different sets of subjects[41,42]. Prediction errors might be easily increased by selecting more stable subjects and calibrating the method to predict little to no changes. For a fair comparison, we re-run our prediction method in the same setting as the TADPOLE challenge and compare our predictions with the 56 competing methods[43]. We estimated another AD Course Map using the TADPOLE training set only and predict future progression of the ADAS-Cog and ventricular volume for the TADPOLE test set. The prediction error of the ADAS-Cog is of $3.70 \pm 0.26(std)$. It is 21% smaller than the best competing method showing an error of 4.70. The constant prediction yields an MAE of 7.077 ± 0.352. Prediction error of the ventricular volume is of $0.412 \pm 0.056(std)$, which is the same as the best performing method.

AD Course Map simulates cohorts of virtual patients. We use AD Course Map to simulate entirely synthetic patients developing Alzheimer's disease. After model calibration, we estimate the empirical posterior distribution of the subject-specific parameters. We sample then random parameters from this distribution. These parameters are used to transform the reference trajectories into a series of trajectories reproducing the heterogeneity seen in the original population. We pick randomly an arbitrary number of time-points in these trajectories to create a synthetic longitudinal data set. We perform this procedure independently for men and women (see Methods).

To validate such simulations, we aim to replicate the original ADNI data set by picking the same number of time-points and at the same frequency as in the ADNI data set, and by simulating men and women subjects with the same sex ratio. We then compute in the simulated data the regional SUVR, cortical thickness, hippocampus volume and neuro-psychological assessments. We superimpose them with the distributions of the original and reconstructed data.

The superimposition of the distributions in Figure 5 shows that the simulated data closely replicate the reconstructed data for all modalities. For the hippocampus volume, the simulated data have the same bias than the reconstructed data in comparison to the real data. This fact is expected as the simulation reproduces the variability learned by the model. This experiment shows that AD Course Map can be used as a simulator of subjects developing Alzheimer's disease, which replicates the heterogeneity in disease progression of the studied population.
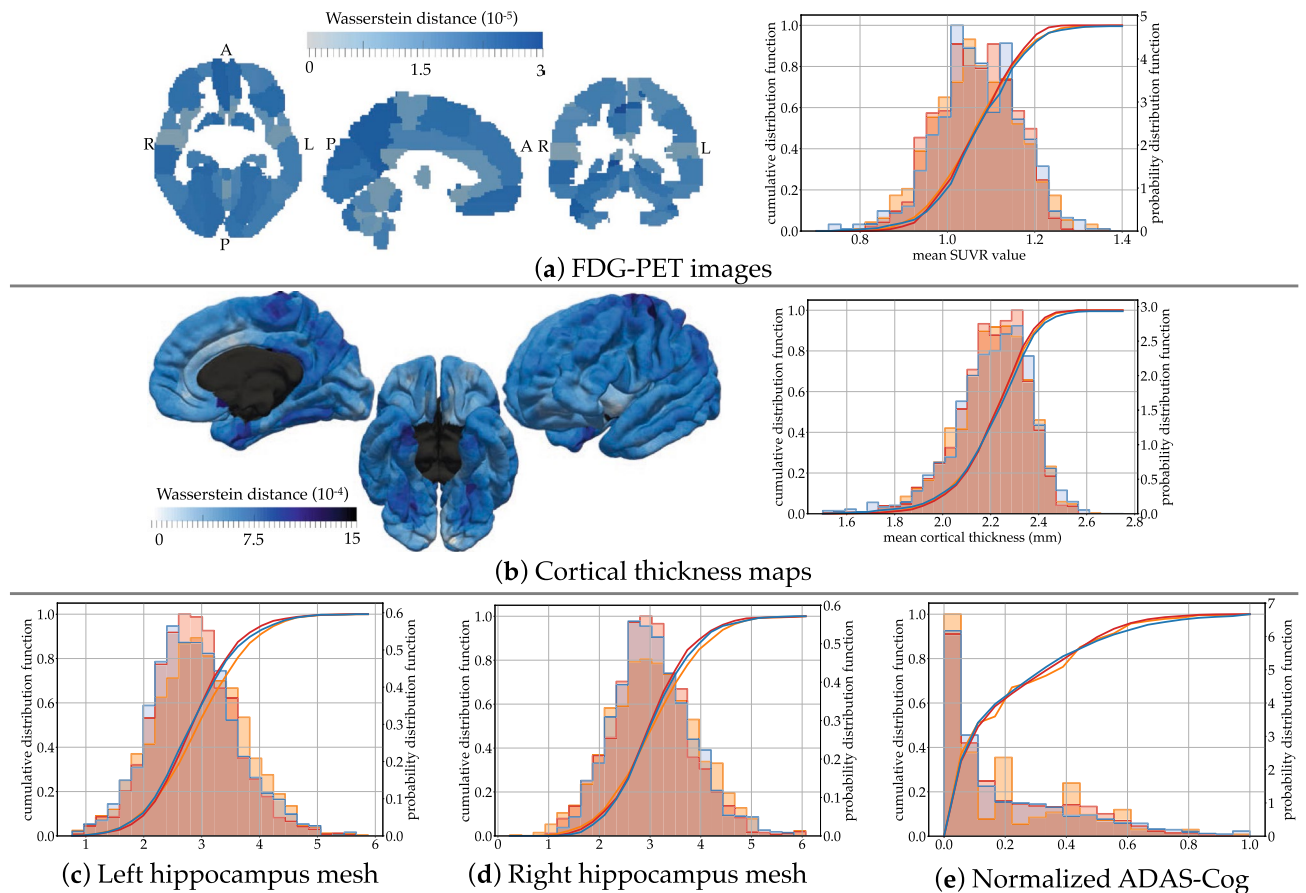
(a) FDG-PET images

(b) Cortical thickness maps

(c) Left hippocampus mesh **(d)** Right hippocampus mesh **(e)** Normalized ADAS-Cog

Figure 5. Statistics of the virtual cohort. Superimposition of empirical distributions for simulated data (blue), reconstructed errors (red, as in Fig. 3) and real data (orange). Images were obtained using the freely available software FSLeyes v.0.22.6 (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLeyes) and Paraview v.5.2.0 (www.paraview.org).

We have released a virtual cohort of one million of subjects developing AD that is freely available at www.digital-brain.org.

This simulator can be used to arbitrarily increase the number of subjects, number of visits and visits frequency in the training cohort. These data augmentation and resampling techniques are essential to improve the performance of machine learning algorithms. It can also be used to create large validation sets to better evaluate the generalisation errors of such algorithms.

Such a simulator can be seen also as an anonymous replicate of the training cohort, which can therefore be transferred and shared without regulatory constraints in lieu of the data itself. In this way, they allow the comparison and combination of multiple cohorts that would otherwise be very difficult to aggregate. They can thus be used to detect the respective biases of these cohorts, and possibly correct them by simulating patients with a re-balanced disease stage distribution, sex ratio, or ratio of APOE-ε4 carriers for instance. The identification of such biases is essential because they are then found in the predictive systems trained on these cohorts.

Discussion

AD Course Map charts the possible trajectories of progression to AD. Few and interpretable parameters adjust a reference scenario of progression to the data of a subject. They allowed us to highlight strong sexual dimorphism and possible compensatory mechanisms at play during AD progression.

The personalized subject's trajectory predicts the future image and clinical data at any time-point in the future. Such predictions are competitive with the state-of-the-art machine learning methods when tested on the same population. The prediction errors are of the order of the measurement uncertainty.

Alternative methods are designed to more directly inform about possible biological mechanisms, such as the propagation of misfolded proteins in the brain[44,45]. Nevertheless, our approach is not specific to certain types of data, and may be applied easily to other therapeutic areas.

One key open question remains the potential use of these disease modeling techniques in clinical decision support systems. The promise of AD Course Map is to target the right patient at the right time. Whether such selection tools eventually lead to more powered clinical trials still needs to be shown.
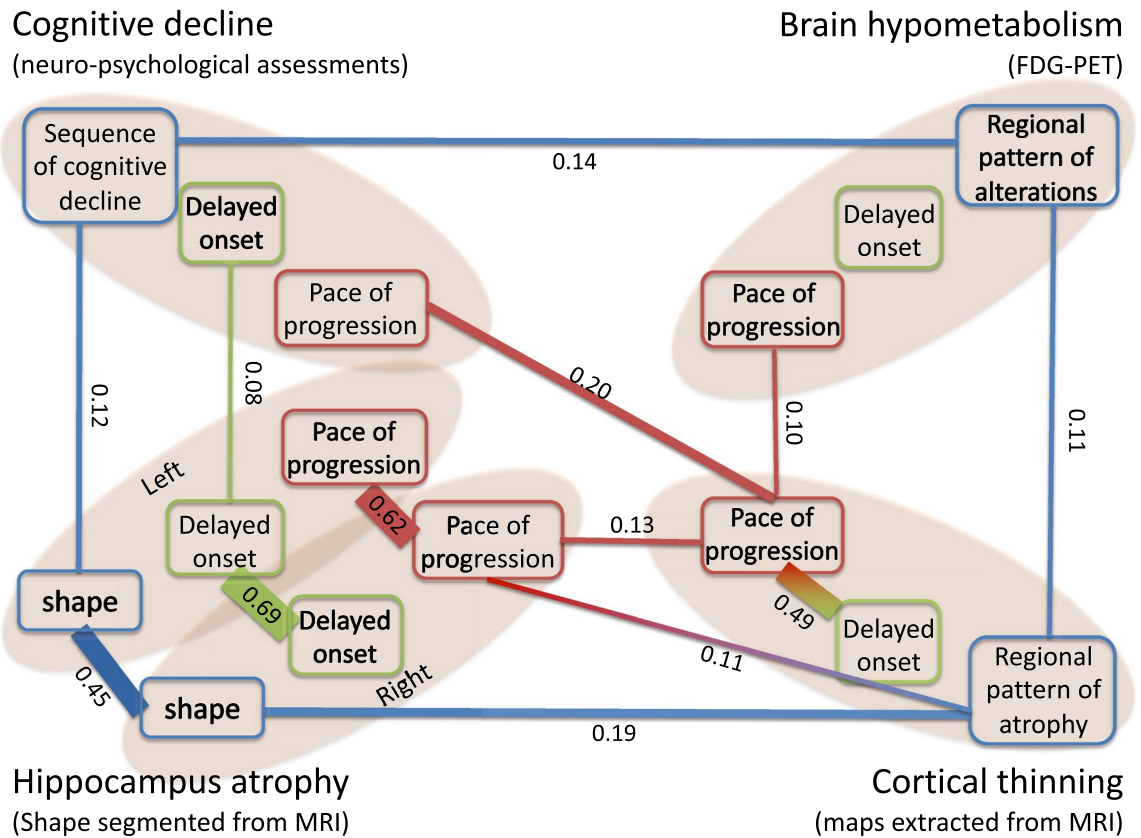
Figure 6. Graph of conditional correlations among individual parameters. An edge is shown between two parameters if there is a significant correlation between them given all other parameters. The width of the edge is proportional to the value of the conditional correlation, which is also reported on the edge. The color of the parameter denotes its type and its position the modality. The image was obtained using the software Microsoft PowerPoint v.16.43.

Methods

Data set.   We use data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) to reproduce the natural history of the disease from the pre-clinical to the clinical stage. We thus selected the 322 subjects in the ADNI, ADNI-GO and ADNI-2 databases who were included as cognitively normal or with mild cognitive impairments and who had a confirmed clinical diagnosis of Alzheimer's disease at a later time-point in the study. Definitions of these categories are as in the ADNI protocol.

Whenever available, we use at each visit:

- regional measurements of standard uptake value ratio (SUVR) of fluorodeoxyglucose (FDG)-positron emission tomography (PET) to build models of hypometabolism across brain regions,
- maps of cortical thickness defined on a mesh of the cortex and extracted from T1-weighted Magnetic Resonance Images (MRI) to build models of cortical thinning,
- surface meshes of the hippocampus of both hemispheres segmented also from T1-weighted MRI to build models of hippocampal atrophy, and
- scores of the Mini-Mental State Examination[46] (MMSE) and Alzheimer's Disease Assessment Scale - Cognitive Subscale with 13 items[47,48] (ADAS-Cog), the latter being divided into four sub-scores assessing memory, language, concentration and praxis, to build models of cognitive decline,

which amounts to 687 visits with PET images, 1993 visits with MRI data and 1235 visits with neuro-psychological assessments (See Supplementary Table S3 for summary statistics).

For each subject, we used the following additional data: age at each visit, sex, marital status, educational level, Apolipoprotein E (ApoE) polymorphism, and presence of amyloidosis. More precisely, we define:

- marital status as: married versus non-married meaning widowed, divorced, or never married;
- educational level as the number of years of education;
- ApoE-ε4 carriership as the presence of at least one allele ε4 of the ApoE gene;
- Amyloid status as positive if one of these conditions was met at one visit at least:

– a Standard Updake Value ratio (SUVR), normalised by the entire cerebellum, greater than 1.1 in a PET image acquired with Florbetapir (AV-45) compound[49,50];
– an average SUVR, normalised by the cerebellum, greater than 1.47 in a PET image with a Pittsburgh compound B (PiB)[50];
– a level of beta amyloid 1-42 (Aβ42) (measured with the Roche Elecsys assays (see http://www.adni.loni.usc.edu/new-csf-a%CE%B21-42-t-tau-and-p-tau181-biomarkers-results-from-adni-biomarker-core-using-elecsys/) in the cerebrospinal fluid (CSF) lower than 1098 pg/mL[51];

unknown if no values of CSF biomarkers and no AV45 or PiB PET images were available at any visit in the ADNI-merge file; and negative otherwise.

Data pre-processing. MMSE total score, ADAS-cog total score and sub-scores were all normalized to a scale of zero to one, such that normal values correspond to zero and the maximum pathological value to one. Sub-scores were obtained by adding the following items of the ADAS-Cog questionnaire: items 1, 4, 7, 8 and 9 for memory, items 2, 5, 10, 11 and 12 for language, items 3 and 6 for praxis, and item 13 for concentration.

Regional FDG-PET SUVR were extracted using the second version of the Automated Anatomical Atlas[52,53] (see http://www.gin.cnrs.fr/fr/outils/aal-aal2/) (AAL2) with 120 regions covering the cortex and the main subcortical structures, using the open-source community software Clinica[54] (see http://clinica.run/doc/Pipelines/PET_Volume). The software performs intra-subject registration of the FDG-PET image into the space of the subject's T1-weighted MRI image using Statistical Parametric Mapping[55] (see www.fil.ion.ucl.ac.uk/spm/ (SPM) software (version 12)). The PET image is then spatially normalised into MNI space using DARTEL deformation model of SPM, and its intensities normalised using the average uptake value in the pons as reference region. The SUVR map is obtained by averaging resulting intensities in each region of the atlas[56].

The MRI images were first processed independently with the cross-sectional pipeline of the FreeSurfer[57,58] (see https://surfer.nmr.mgh.harvard.edu software (version 5.3.0)). The longitudinal FressSurfer pipeline is then used to create subject-specific templates from the successive data of each subject and refine image segmentations[59]. These segmented images are used then to extract a cortical thickness map, and a mesh of the left and right hippocampus.

We used the cortical surface mesh projected onto the average space called FSaverage with 163,842 vertices. For dimensionality reduction purposes, we then

- inflate the FSAverage mesh to a sphere using FreeSurfer, on which 3658 vertices (called patch-nodes) are selected to map the whole sphere uniformly,
- associate each vertex to its closest patch-node, resulting in a parcellation of the cortical mesh into 3658 patches that are uniformly distributed over the surface, where a patch contains on average 44 vertices,
- compute the average value of the cortical thickness in each patch.

We also aligned the skull-stripped images with an affine 12-degrees-of-freedom transformation onto the Colin27 template brain (see http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27, using the FSL 5.0 softwarecitewoolrich2009bayesian (see https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/). Mesh representations of the geometry of the left and right hippocampus result from the following steps:

- the volumetric segmentations of the hippocampi obtained by FreeSurfer are transformed into meshes using the aseg2srf software(see https://brainder.org (version of July 2009)),
- the resulting meshes are decimated by a 88% factor using Paraview, 5.4.1[60] (see www.paraview.org),
- then aligned using the previously-computed global affine transformation estimated with the FSL software,
- residual pose differences among subjects are then removed by rigidly aligning the meshes from the baseline image of each subject to the corresponding hippocampus mesh in the Colin27 atlas image, this transformation with 6 degrees of freedom being computed with the GMMReg software[61] (see https://github.com/bing-jian/gmmreg (version of July 2008)),
- the same transformation is eventually used to align the meshes from the follow-up images of the same subject.

Disease course mapping. Disease course mapping models changes in image and clinical data during disease progression. Its core aspect is the construction a long-term normative scenario of progression by normalizing and mapping subject's data to different portions of a common trajectory. This mapping includes notably the estimation of an Alzheimer Age at each visit of each patient.

We consider first the simple case where one observes only two biomarkers at each visit of each patient. As illustrated in the Figure 1 (top row), the idealized model of progression is mapped to the individual data of a subject using three parameters:

- time-shift τ_i, the temporal onset of disease progression for all biomarkers in subject i; τ_i is negative for earlier-than-average progression and positive for later-than-average progression,
- acceleration α_i, the change in the rate of progression of subject i; $\alpha_i = 1$ means that, for that subject, all the biomarker trajectories match the idealized curves after adjusting the timing with τ_i and ω_i. When $\alpha_i > 1$, it takes less time for the same changes to occur, e.g. $\alpha_i = 2$ means that it will twice as fast as average for that subject to undergo a given change in a biomarker trajectory,
- intermarker spacing ω_{ik}, the temporal ordering of biomarker k with respect to other biomarkers for patient i. The sign of ω_{ik} indicates whether a single feature in a particular subject exhibits earlier or later changes

relative to other features (after normalizing the age at onset with τ_i and pace of progression with α_i: $\omega_{ik} < 0$ means earlier degradation, and $\omega_{ik} > 0$ means later.

For the model to be identifiable, the effect of the intermarker spacing on the curves should not overlap with the time-shift τ_i. If the two logistic curves are parallel, one intermarker spacing should be the opposite to the other one. If the two logistic curves are not parallel, a weighted sum of the inter-marker spacing should sum up to zero, where the weights depend on the parameters of the logistics. The consequence is that for N biomarkers, there is only $N - 1$ independent intermarker spacing[11,12].

To go further, we now plot one biomarker versus the other one in Figure 1 (bottom row). The logistic progression of the two biomarkers forms a curve in the unit square starting at point of coordinate $(0, 0)$ when both biomarkers are normal and ending at point of coordinate $(1, 1)$ when both biomarkers reach their maximal value. One point represents the two biomarkers of one patient at one time-point. Time is implicit in this representation.

The effect of the time-shift τ and acceleration α is to slide the points along the curve so that the actual age of the subject at each visit moves to the corresponding Alzheimer Age (i.e. the time index of the points closest to the curve). These parameters change the parametric representation of the curve, not the curve itself. By contrast, the effect of the interspacing marker is to translate the subject's curve in the direction given by the vector $\omega = (\omega_1, \omega_2)$ to fit the reference population curve. We impose the vector ω to be orthogonal to the velocity of the curve to ensure that the effect of the intermarker spacing does not overlap with that of τ and α. This condition amounts to imposing that a weighted average of the coordinates of the intermarker spacing equal zero.

This representation of the method permits now its generalisation to a large class of data. In the previous example, the unit square is a particular case of a Riemannian manifold, a mathematical space that generalise the usual Euclidean plane. The central curve is a geodesic on this manifold, namely a curve of minimal energy that is fully determined by its velocity at a given point and time. The change in the parametric representation of the curve with the time-shift τ and acceleration α applies to any types of curves. The translation in the direction of ω is a generalisation of the usual translation that we have introduced and called "Exp-parallelisation"[11,12].

Therefore, the exact same construction applies for any data which may be represented on a Riemannian manifold. One must only specify the Riemmanian manifold, in particular its metric which determines the parametric form of the geodesics. Then, the normative scenario of progression is assumed to be a geodesics in this manifold. Subjects are assumed to follow a piece of curve that derive from the normative curve by a change in the parametric representation and an Exp-parallelisation. This construction defines a spatiotemporal coordinate system around the reference curve that is used to position the subjects trajectories.

We consider cortical thickness measurements and images as a series of biomarkers (one per region, voxel or vertex). We define the Riemannian metric such that each biomarker follow a straight line and the intercept and slopes of the straight lines vary smoothly across neighbourhing regions, voxels or vertices[62,63]. We consider changes in the shape of the hippocampus are due to a smooth deformation of the surface called diffeomorphisms[64,65].

The central geodesics, namely the idealized set of logistics in the two biomarkers example, serves as a reference pattern of progression. Our fitting to this curve requires simultaneous estimation of the individual parameters (τ_i, α_i, and ω_{ik} of each subject in a longitudinal setting. Each subject has a single τ, a single α and a multivariate ω (one per biomarker in the biomarker example, in other examples the size of ω is the dimension of the considered Riemmanian manofold). We estimate the parameters of the geodesics, namely the shape and position of the logistic curves in the biomarker example, so that mean value of the τ_i, $\log(\alpha_i)$ and each of the ω_{ij} across all subjects is zero. The resulting trajectory represents the progression of the average subject in the studied population. We iteratively perform joint estimations of the population and individual parameters to minimize the residual variance.

The disease course map results from the calibration of the method on a longitudinal data with at least two time-points per subject. The disease course map includes the estimated normative scenario of change and the empirical distribution of the individual parameters τ, α and ω.

Once the disease course map is calibrated, one can position new subjects in the map and predict their future progression. We use the baseline data of the subjects, with additional follow-up if available, to estimate the individual parameters for the subjects and transform the normative scenario of changes to match empirical data (see Figure 1). One the one hand, the individual parameters position subjects among themselves. One may then test for example if some genetic factors favor early or late disease onset, fast or slow progression, or specific patterns of progression such as a specific ordering in the decline of cognitive functions. One the other hand, the personalized scenario provides a prediction of the future values of the subject's data at any time-point in the future. Eventually, one may simulate random transformation of the reference scenario to synthesise virtual patients developing the disease.

A non linear mixed-effects model. All in one, this construction presented above defines a non-linear mixed-effects statistical model. We denote $\gamma_0(t)$ the population geodesic curve where t is the Alzheimer age, $\eta^{w_i}[\gamma_0](t)$ the parallel shift of the population curve in the direction w_i, and $\psi_i(t) = \alpha_i(t - t_0 - \tau_i) + t_0$ the time-reparameterisation function, called time-warp, determined by the time-shift τ_i and acceleration α_i. It maps the actual age of the subject to its Alzheimer age. The value of $\gamma_0(t)$ at time t results from the integration of a second-order ordinary differential equation. It is therefore a function of its initial conditions: an initial point p_0 and a velocity v_0 at a given time-point t_0.

The j-th observation of the i-th subject, denoted y_{ij} acquired at actual age t_{ij} is then assumed to be derived from the population curve by $y_{ij} = \eta^{w_i}[\gamma_0](\psi_i(t_{ij})) + \varepsilon_{ij}$ for the ε_{ij} is a random noise.

The model may be written in short as $y_{ij} = f(\theta, z_i, t_{ij}) + \varepsilon_{ij}$, for f a non-linear function that is specific to each data type, θ the vector containing the fixed-effects p_0, v_0, t_0, the variance of the random-effects and the variance of the noise, and z_i the vector of random effects: acceleration factors, time-shifts and space-shifts. We add priors on the coordinates of the vector θ in a Bayesian setting. When t is varied, the curve $f(\theta, z_i, t)$ represents the subject-specific trajectory at any time t.

We now consider four statistical tasks:

- **calibration:** given the longitudinal data set $\{y_{ij}, t_{ij}\}_{i=1,\dots,N, j=1,\dots,N_i}$, we find the value of parameters θ that maximises the joint likelihood $p(\{y_{ij}\}_{ij}, \theta) = p(\{y_{ij}\}_{ij} | \theta) p(\theta)$. The optimal value $\hat{\theta}$ fully specifies the disease course map;
- **personalisation:** for the optimal value of the parameter $\hat{\theta}$, we personalise the model to the repeated data of a given subject (either a training subject, or a test subject in a cross-validation setting) $\{y_{test,j}, t_{test,j}\}_{j=1,\dots,N_{test}}$ by finding the optimal value of the random-effect $\hat{z}$ that maximises the conditional likelihood $p(\{y_{test,j}\}_j, z | \hat{\theta})$. The resulting $f(\hat{\theta}, \hat{z}, t_{test,j})$ is called the **reconstruction** of the data $y_{test,j}$ and its difference with the true data $y_{test,j}$ is called the reconstruction error;
- **prediction:** given a test subject with N_{test} observations, we personalize the model using only the first N_{past} ($< N_{test}$) observations to estimate $\hat{z}$, and then predict the future data after N_{past} by extrapolating the trajectory $f(\hat{\theta}, \hat{z}, t_{test,j})$, and measure the prediction error between the predicted and true (hidden) data.
- **simulation:** for the optimal value of the parameter $\hat{\theta}$, we simulate random-effects z and generate synthetic data y at any user-defined time-point t by computing $y = f(\hat{\theta}, z, t)$ and adding noise.

We use a stochastic approximation of the Expectation-Minimisation algorithm[66–68] for calibration, gradient-descent based method or Powell's method for personalisation, and kernel density estimation together with dimension reduction for simulation.

We computed Student t-test between cofactors and estimates of the random effects using the False discovery rate method for the correction for multiple comparisons. A graph of conditional correlations is obtained using an exploration and selection method adapted in the high-dimensional low sample size setting[69].

Estimation of measurement uncertainty. In the ADNI protocol[70,71], most MRI sessions consist of a pair of test and re-test MRI, namely two scans performed on the same day one immediately after the other one. For 1,841 out of 1,993 MRI sessions, we measure therefore the differences between the MRI derived data (hippocampus meshes and cortical thickness maps) when using the test or the re-test image in the processing pipeline describe above. These differences give an empirical distribution of the noise due to variations in image acquisition and processing.

For PET derived data, we use the baseline and follow-up scans of stable cognitively normal and amyloid negative subjects in ADNI, as a proxy to test / re-test data (125 subjects, 244 visits with a follow-up time of 18 months).

Test / re-test studies have shown a that the MMSE, which scales from 0 to 30, is subject to a difference between two sessions, whose standard deviation ranges from 1.3 for a one-month interval[39] up to 1.82 for a 1.5 year long interval[72], thus representing a standard deviation of 4.3 to 6%. Another study[73] measured the former ADAS-Cog that scales between 0 and 70 three times at a 2-week interval, with an agreement between raters. The inter-ratter standard deviation is of 9.64 between the first and second test, and of 6.79 between the second and third test. The intra-rater standard deviation is of 8.16 between the first and third visit. This corresponds to a standard deviation ranging from 9.7% to 13.8%. On average, we consider such neuro-psychological assessments to have a zero-mean Gaussian distribution of noise with standard deviation of 7%.

References

1. Fitzmaurice, G., Laird, N. & Ware, J. *Applied longitudinal analysis* 2nd edn. (John Wiley and sons, 2011).
2. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G. & Ayache, N. Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2009, vol. 5761 of Lecture Notes in Computer Science* (eds Yang, G.-Z. et al.) 297–304 (Springer, Berlin, 2009).
3. Durrleman, S. et al. Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data. *Int. J. Comput. Vis.* **103**, 22–59. https://doi.org/10.1007/s11263-012-0592-x (2013).
4. Donohue, M. C. et al. Estimating long-term multivariate progression from short-term data. *Alzheimer's & Dementia J. Alzheimer's Assoc.* **10**, S400-410. https://doi.org/10.1016/j.jalz.2013.10.003 (2014).
5. Taddé, B. O., Jacqmin-Gadda, H., Dartigues, J., Commenges, D. & Proust-Lima, C. Dynamic modeling of multivariate dimensions and their temporal relationships using latent processes: application to alzheimer's disease. *Biometrics* **76**, 886–899 (2020).
6. Bilgel, M. et al. A multivariate nonlinear mixed effects model for longitudinal image analysis: application to amyloid imaging. *NeuroImage* **134**, 658–670 (2016).
7. Marinescu, R. et al. Dive: a spatiotemporal progression model of brain pathology in neurodegenerative disorders. *NeuroImage* **192**, 166–177 (2019).
8. Bilgel, M., Jedynak, B. M. & Initiative, A. D. N. Predicting time to dementia using a quantitative template of disease progression. *Alzheimer's & Dementia Diagn. Assess. Dis. Monit.* **11**, 205–215 (2019).
9. Fonteijn, H. M. et al. An event-based model for disease progression and its application in familial, alzheimer's disease and huntington's disease. *NeuroImage* **60**, 1880–1889 (2012).
10. Archetti, D. et al. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in alzheimer's disease. *Neuroimage Clin.* https://doi.org/10.1016/j.nicl.2019.101954 (2019).

11. Schiratti, J.-B., Allassonnière, S., Colliot, O. & Durrleman, S. Learning spatiotemporal trajectories from manifold-valued longitudinal data. In *NIPS 28* (eds Cortes, C. *et al.*) 2404–2412 (Curran Associates Inc, 2015).

12. Schiratti, J.-B., Allassonnière, S., Colliot, O. & Durrleman, S. A bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. *J. Mach. Learn. Res.* **18**, 4840–4872 (2017).

13. Skinner, J. *et al.* The alzheimer's disease assessment scale-cognitive-plus (adas-cog-plus): an expansion of the adas-cog to improve responsiveness in mci. *Brain Imaging Behav.* **6**, 489–501 (2012).

14. Raghavan, N. *et al.* The adas-cog revisited: novel composite scales based on adas-cog to improve efficiency in mci and early ad trials. *Alzheimer's & Dementia* **9**, S21–S31 (2013).

15. Mosconi, L. Brain glucose metabolism in the early and specific diagnosis of alzheimer's disease. *Eur. J. Nucl. Med. Mol. Imaging* **32**, 486–510 (2005).

16. Chen, K. *et al.* Twelve-month metabolic declines in probable alzheimer's disease and amnestic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: findings from the alzheimer's disease neuroimaging initiative. *Neuroimage* **51**, 654–664 (2010).

17. Pagani, M. *et al.* Early identification of mci converting to ad: a fdg pet study. *Eur. J. Nucl. Med. Mol. Imaging* **44**, 2042–2052 (2017).

18. Drzezga, A. *et al.* Cerebral metabolic changes accompanying conversion of mild cognitive impairment into alzheimer's disease: a pet follow-up study. *Eur. J. Nucl. Med. Mol. Imaging* **30**, 1104–1113 (2003).

19. Mosconi, L. *et al.* Hippocampal hypometabolism predicts cognitive decline from normal aging. *Neurobiol. Aging* **29**, 676–692 (2008).

20. Hyman, B. T., Van Hoesen, G. W., Damasio, A. R. & Barnes, C. L. Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science* **225**, 1168–1170 (1984).

21. Gómez-Isla, T. *et al.* Profound loss of layer ii entorhinal cortex neurons occurs in very mild alzheimer's disease. *J. Neurosci.* **16**, 4491–4500 (1996).

22. Greene, S. J. *et al.* Subregions of the inferior parietal lobule are affected in the progression to alzheimer's disease. *Neurobiol. Aging* **31**, 1304–1311 (2010).

23. Chan, D. *et al.* Patterns of temporal lobe atrophy in semantic dementia and alzheimer's disease. *Ann. Neurol.* **49**, 433–442 (2001).

24. Jacobs, H. I., Van Boxtel, M. P., Jolles, J., Verhey, F. R. & Uylings, H. B. Parietal cortex matters in alzheimer's disease: an overview of structural, functional and metabolic findings. *Neurosci. Biobehav. Rev.* **36**, 297–309 (2012).

25. Suva, D. *et al.* Primary motor cortex involvement in alzheimer disease. *J. Neuropathol. Exp. Neurol.* **58**, 1125–1134 (1999).

26. Durrleman, S. *et al.* Morphometry of anatomical shape complexes with dense deformations and sparse parameters. *NeuroImage* (2014).

27. Shokouhi, S. *et al.* Longitudinal progression of cognitive decline correlates with changes in the spatial pattern of brain 18f-fdg pet. *J. Nuclear Med.* **54**, 1564–1569 (2013).

28. Ossenkoppele, R. *et al.* Longitudinal imaging of alzheimer pathology using [11c]pib, [18f]fddnp and [18f]fdg pet. *Eur. J. Nucl. Med. Mol. Imaging* **39**, 990–1000 (2012).

29. Ishibashi, K. *et al.* Longitudinal effects of aging on 18f-fdg distribution in cognitively normal elderly individuals. *Sci. Rep.* , https://doi.org/10.1038/s41598-018-29937-y (2018).

30. Knopman, D. S. *et al.* 18f-fluorodeoxyglucose positron emission tomography, aging, and apolipoprotein e genotype in cognitively normal persons. *Neurobiol. Aging* **35**, 2096–2106 (2014).

31. Jack, C. R. *et al.* Age, sex, and apoe ε4 effects on memory, brain structure, and β-amyloid across the adult life span. *JAMA Neurol.* **72**, 511–519 (2015).

32. Raj, A., Kuceyeski, A. & Weiner, M. A network diffusion model of disease progression in dementia. *Neuron* **73**, 204–1215 (2012).

33. Bhargava, D., Weiner, M. F., Hynan, L. S., Diaz-Arrastia, R. & Lipton, A. M. Vascular disease and risk factors, rate of progression, and survival in alzheimer's disease. *J. Geriatr. Psychiatry Neurol.* **19**, 78–82. https://doi.org/10.1177/0891988706286505 (2006).

34. Wattmo, C. & Wallin, A. K. Early- versus late-onset alzheimer's disease in clinical practice: cognitive and global outcomes over 3 years. *Alzheimer's Res. Therapy* **9**, 70. https://doi.org/10.1186/s13195-017-0294-2 (2017).

35. Burns, A., Jacoby, R. & Levy, R. Progression of cognitive impairment in alzheimer's disease. *J. Am. Geriatr. Soc.* **39**, 39–45. https://doi.org/10.1111/j.1532-5415.1991.tb05904.x (1991).

36. Barocco, F. *et al.* The progression of alzheimer's disease: Are fast decliners really fast? a four-year follow-up. *Journal of Alzheimer's disease?: JAD* **57**, 775–786. https://doi.org/10.3233/JAD-161264 (2017).

37. Albert, M. *et al.* The diagnosis of mild cognitive impairment 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.* **7**, 270–279. https://doi.org/10.1016/j.jalz.2011.03.008 (2011).

38. Budd Haeberlein, S. *et al.* Clinical development of aducanumab, an anti-aβ human monoclonal antibody being investigated for the treatment of early alzheimer's disease. *J Prev Alzheimers Dis.* **4**, 255–263. https://doi.org/10.14283/jpad.2017.39 (2017).

39. Clark, C. M. *et al.* Variability in annual mini-mental state examination score in patients with probable alzheimer disease: a clinical perspective of data from the consortium to establish a registry for alzheimer's disease. *Arch. Neurol.* **56**, 857–862 (1999).

40. Leung, K. *et al.* Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. *Neuroimage* **50**, 516–523 (2020).

41. Huang, L., Jin, Y., Gao, Y., Thung, K. H. & Shen, D. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiol. Aging* **46**, 180–191 (2016).

42. Iddi, S. *et al.* Predicting the course of alzheimer's progression. *Brain Inf.* **6**, 6. https://doi.org/10.1186/s40708-019-0099-0 (2019).

43. Marinescu, R. V. *et al.* Tadpole challenge: Prediction of longitudinal evolution in alzheimer's disease (2018). arXiv:1805.03909.

44. Iturria-Medina, Y., Sotero, R., Toussaint, P., Evans, A. & the Alzheimer's Disease Neuroimaging Initiative. Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders. *PLoS Comput. Biol.* **10** https://doi.org/10.1371/journal.pcbi.1003956 **(2014).**

45. Garbarino, S., Lorenzi, M. & the Alzheimer's Disease Neuroimaging Initiative. Modeling and inference of spatio-temporal protein dynamics across brain networks. In Chung, A., Gee, J., Yushkevich, P. & Bao, S. (eds.) *Information Processing in Medical Imaging. IPMI 2019*, vol. 11492 of *Lecture Notes in Computer Science*, (Springer, Cham, 2019).

46. Folstein, M. F., Folstein, S. E. & McHugh, P. R. "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).

47. Rosen, W. G., Mohs, R. C. & Davis, K. L. A new rating scale for alzheimer's disease. *Am. J. Psychiatry* (1984).

48. Mohs, R. C. *et al.* Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the alzheimer's disease assessment scale that broaden its scope. *Alzheimer Dis. Assoc. Disord.* (1997).

49. Clark, C. M. *et al.* Cerebral pet with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study. *Lancet Neurol.* **11**, 669–678 (2012).

50. Landau, S. M. *et al.* Amyloid-β imaging with pittsburgh compound b and florbetapir: comparing radiotracers and quantification methods. *J. Nuclear Med. Offic. Publ. Soc. Nuclear Med.* **54**, 70 (2013).

51. Schindler, S. E. *et al.* Cerebrospinal fluid biomarkers measured by elecsys assays compared to amyloid imaging. *Alzheimer's & Dementia* (2018).

52. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage* **15**, 273–289 (2002).

53. Rolls, E. T., Joliot, M. & Tzourio-Mazoyer, N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *Neuroimage* **122**, 1–5 (2015).
54. Routier, A. *et al.* Clinica: an open source software platform for reproducible clinical neuroscience studies. *Annual meeting of the Organization for Human Brain Mapping-OHBM*, (2018).
55. Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J. & Nichols, T. E. *Statistical parametric mapping: the analysis of functional brain images* (Elsevier, 2011).
56. Samper-González, J. *et al.* Reproducible evaluation of classification methods in alzheimer's disease: framework and application to mri and pet data. *bioRxiv* 274324 (2018).
57. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Nat. Acad. Sci.* **97**, 11050–11055 (2000).
58. Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
59. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* **61**, 1402–1418 (2012).
60. Ahrens, J., Geveci, B. & Law, C. Paraview: An end-user tool for large data visualization. *Visualiz. Handb.* **717**, (2005).
61. Jian, B. & Vemuri, B. C. Robust point set registration using gaussian mixture models. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 1633–1645 (2011).
62. Koval, I. *et al.* Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 451–459 (Springer, 2017).
63. Koval, I. *et al.* Spatiotemporal propagation of the cortical atrophy: Population and individual patterns. *Frontiers Neurol.* **9**, 235. https://doi.org/10.3389/fneur.2018.00235 (2018).
64. Bône, A., Colliot, O. & Durrleman, S. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* **9271–9280**, (2018).
65. Bône, A., Colliot, O. & Durrleman, S. Learning the spatiotemporal variability in longitudinal shape data sets. *Int. J. Comput. Vis.* **128**, 2873–2896. https://doi.org/10.1007/s11263-020-01343-w (2020).
66. Kuhn, E. & Lavielle, M. Coupling a stochastic approximation version of em with an mcmc procedure. *ESAIM: Probab. Stat.* **8**, 115–131 (2004).
67. Allassonnière, S., Kuhn, E. & Trouvé, A. Construction of bayesian deformable models via a stochastic approximation algorithm: a convergence study. *Bernoulli* **16**, 641–678 (2010).
68. Allassonnière, S., Durrleman, S. & Kuhn, E. Bayesian mixed effect atlas estimation with a diffeomorphic deformation model. *SIAM J. Imaging Sci.* **8**, 1367–1395 (2015).
69. Lartigue, T. *et al.* Gaussian graphical model exploration and selection in high dimension low sample size setting. *IEEE Trans. Pattern Anal. Mach. Intell.*, https://doi.org/10.1109/TPAMI.2020.2980542 (5555).
70. Jack, C. R. Jr. *et al.* The alzheimer's disease neuroimaging initiative (adni): Mri methods. *J. Magn. Reson. Imaging Offic. J. Int. Soc. Magn. Reson. Med.* **27**, 685–691 (2008).
71. Jack, C. R. Jr. *et al.* Update on the magnetic resonance imaging core of the alzheimer's disease neuroimaging initiative. *Alzheimer's & Dementia* **6**, 212–220 (2010).
72. Hensel, A., Angermeyer, M. C. & Riedel-Heller, S. G. Measuring cognitive change in older adults: reliable change indices for the mmse. *J. Neurol. Neurosurg. Psychiatry* (2007).
73. Standish, T. I. *et al.* Improved reliability of the standardized alzheimer's disease assessment scale (sadas) compared with the alzheimer's disease assessment scale (adas). *J. Am. Geriatr. Soc.* **44**, 712–716 (1996).

Acknowledgements

Author contributions

I.K, A.B., M.L., T.L. performed the research, S.B., A.M., J.S. managed and pre-processed data, N.B., B.C., A.B., S.E., O.C., S.A., S.D. designed the research, I.K., A.B., T.L., N.B., S.E., O.C., S.A, S.D wrote the paper. All authors reviewed the manuscript.

Competing interests

Additional information

Supplementary Information is available for this paper at https://doi.org/10.1038/s41598-021-87434-1.

Correspondence and requests for materials should be addressed to S.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.