Simulating the outcome of amyloid treatments in Alzheimer's Disease from multi-modal imaging and clinical data

Clément Abi Nader[a,*], Nicholas Ayache[a], Giovanni B. Frisoni[b], Philippe Robert[c], Marco Lorenzi[a], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a] *Université Côte d'Azur, Inria Sophia Antipolis, Epione Research Project, France.*
[b] *Memory Clinic and LANVIE-Laboratory of Neuroimaging of Aging, Hospitals and University of Geneva, Geneva, Switzerland*
[c] *Université Côte d'Azur, CoBTeK lab, MNC3 program, France.*

Abstract

Recent failures of clinical trials in Alzheimer's Disease underline the critical importance of identifying optimal intervention time to maximize cognitive benefit. While several models of disease progression have been proposed, we still lack quantitative approaches simulating the effect of treatment strategies on the clinical evolution. In this work, we present a data-driven method to model dynamical relationships between imaging and clinical biomarkers. Our approach allows simulating intervention at any stage of the pathology by modulating the progression speed of the biomarkers, and by subsequently assessing the impact on disease evolution. When applied to multi-modal imaging and clinical data from the Alzheimer's Disease Neuroimaging Initiative our method enables to generate hypothetical scenarios of amyloid lowering interventions. Our results show that in a study with 1000 individuals per arm, accumulation should be completely arrested at least 5 years before Alzheimer's dementia diagnosis to lead to statistically powered improvement of clinical endpoints.

1. Introduction

The number of people affected by Alzheimer's Disease (AD) has recently exceeded 46 millions and is expected to double every 20 years [1], thus posing significant healthcare challenges. Yet, while the disease mechanisms remain in large part unknown, there are still no effective pharmacological treatments leading to tangible improvements of patients' clinical progression. One of the main challenges in understanding AD is that its progression goes through a silent asymptomatic phase that can stretch over decades before a clinical diagnosis can be established based on cognitive and behavioral symptoms. To help designing appropriate intervention strategies, hypothetical models of the disease history have been proposed, characterizing the progression by a cascade of morphological and molecular changes affecting the brain, ultimately leading to cognitive impairment [2, 3]. The dominant hypothesis is that disease dynamics along the asymptomatic period are driven by the deposition in the brain of the amyloid β peptide, triggering the so-called "amyloid cascade" [4, 5, 6, 7, 8]. Based on this rationale, clinical trials have been focusing on the development and testing of disease modifiers targeting amyloid β aggregates [9], for example by increasing its clearance or blocking its accumulation. Although the amyloid hypothesis has been recently invigorated by a post-hoc analysis of the aducanumab trial [10], clinical trials failed so far to show efficacy of this kind of treatments, as the clinical primary endpoints were not met [11, 12, 13], or because of unacceptable adverse effects [14]. In the past years, growing consensus emerged about the critical importance of intervention time, and about the need of starting anti-amyloid treatments during the pre-symptomatic stages of the disease [15]. Nevertheless, the design of optimal

*Corresponding author at: Epione Research Project, INRIA Sophia-Antipolis, 2004, route des Lucioles, 06902 Sophia-Antipolis, France, clement.abi-nader@inria.fr.

[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.

intervention strategies is currently not supported by quantitative analysis methods allowing to model and assess the effect of intervention time and dosing [16]. The availability of models of the pathophysiology of AD would entail great potential to test and analyze clinical hypothesis characterizing AD mechanisms, progression, and intervention scenarios.

Within this context, quantitative models of disease progression, referred to as Disease Progression Models (DPMs), have been proposed [17, 18, 19, 20, 21], to quantify the dynamics of the changes affecting the brain during the whole disease span. These models rely on the statistical analysis of large datasets of different data modalities, such as clinical scores, or brain imaging measures derived from Magnetic Resonance Imaging (MRI), Amyloid- and Fluorodeoxyglucose-Positron Emission Tomography (PET) [22, 23, 24]. In general, DPMs estimate a long-term disease evolution from the joint analysis of multivariate time-series acquired on a short-term time-scale. Due to the temporal delay between the disease onset and the appearance of the first symptoms, DPMs rely on the identification of an appropriate temporal reference to describe the long-term disease evolution [25, 26]. These tools are promising approaches for the analysis of clinical trials data, as they allow to represent the longitudinal evolution of multiple biomarkers through a global model of disease progression. Such a model can be subsequently used as a reference in order to stage subjects and quantify their relative progression speed [27, 28, 29]. However, these approaches remain purely descriptive as they don't account for causal relationships among biomarkers. Therefore, they generally don't allow to simulate progression scenarios based on hypothetical intervention strategies, thus providing a limited interpretation of the pathological dynamics. This latter capability is of utmost importance for planning and assessment of disease modifying treatments.

To fill this gap, recent works such as [30, 31] proposed to model AD progression based on specific assumptions on the biochemical processes of pathological protein propagation. These approaches explicitly define biomarkers interactions through the specification of sets of Ordinary Differential Equations (ODEs), and are ideally suited to simulate the effect of drug interventions [32]. However, these methods are mostly based on the arbitrary choices of pre-defined evolution models, which are not inferred from data. This issue was recently addressed by [33], where the authors proposed an hybrid modeling method combining traditional DPMs with dynamical models of AD progression. Still, since this approach requires to design suitable models of protein propagation across brain regions, extending this method to jointly account for spatio-temporal interactions between several processes, such as amyloid propagation, glucose hypometabolism, and brain atrophy, is considerably more complex. Finally, these methods are usually designed to account for imaging data only, which prevents to jointly simulate heterogeneous measures [34], such as image-based biomarkers and clinical outcomes, the latter remaining the reference markers for patients and clinicians.

In this work we present a novel computational model of AD progression allowing to simulate intervention strategies across the history of the disease. The model is here used to quantify the potential effect of amyloid modifiers on the progression of brain atrophy, glucose hypometabolism, and ultimately on the clinical outcomes for different scenarios of intervention. To this end, we model the joint spatio-temporal variation of different modalities along the history of AD by identifying a system of ODEs governing the pathological progression. This latent ODEs system is specified within an interpretable low-dimensional space relating multi-modal information, and combines clinically-inspired constraints with unknown interactions that we wish to estimate. The interpretability of the relationships in the latent space is ensured by mapping each data modality to a specific latent coordinate. The model is formulated within a Bayesian framework, where the latent representation and dynamics are efficiently estimated through stochastic variational inference. To generate hypothetical scenarios of amyloid lowering interventions, we apply our approach to multi-modal imaging and clinical data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our results provide a meaningful quantification of different intervention strategies, compatible with findings previously reported in clinical studies. For example, we estimate that in a study with 100 individuals per arm, statistically powered improvement of clinical endpoints can be obtained by completely arresting amyloid accumulation at least 8 years before Alzheimer's dementia. The minimum intervention time decreases to 5 years for studies based on 1000 individuals per arm.

2. Results

In the following sections, healthy individuals will be denoted as NL stable, subjects with mild cognitive impairment as MCI stable, subjects diagnosed with Alzheimer's dementia as AD, subjects progressing from NL to MCI as

NL converters, and subjects progressing from MCI to AD as MCI converters. Amyloid concentration and glucose metabolism are respectively measured by (18)F-florbetapir Amyloid (AV45)-PET and (18)F-fluorodeoxyglucose (FDG)-PET imaging. Cognitive and functional abilities are assessed by the following neuro-psychological tests: Alzheimer's Disease Assessment Scale (ADAS11), Mini-Mental State Examination (MMSE), Functional Assessment Questionnaire (FAQ), Rey Auditory Verbal Learning Test (RAVLT) immediate, and RAVLT forgetting.

2.1. Study cohort and biomarkers' changes across clinical groups

Our study is based on a cohort of 311 amyloid positive individuals composed of 46 NL stable subjects, 10 NL converters subjects, 106 subjects diagnosed with MCI, 76 MCI converters subjects, and 73 AD patients. The term "amyloid positive" refers to subjects whose amyloid level in the cerebrospinal fluid (CSF) was below the nominal cutoff of 192 pg/ml [35] either at baseline, or during any follow-up visit, and conversion to AD was determined using the last available follow-up information. The length of follow-up varies between subjects and goes from 0 to 6 years. Further information about the data are available on `https://adni.bitbucket.io/reference/`, while details on data acquisition and processing are provided in Section 4.1. We show in Table 1A socio-demographic information for the training cohort across the different clinical groups. Table 1B shows baseline values and annual rates of change across clinical groups for amyloid burden (average normalized AV45 uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), glucose hypometabolism (average normalized FDG uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), for hippocampal and medial temporal lobe volumes, and for the cognitive ability as measured by ADAS11. Compatibly with previously reported results [36, 37], we observe that while regional atrophy, hypometabolism and cognition show increasing rate of change when moving from healthy to pathological conditions, the change of AV45 is maximum in NL stable and MCI stable subjects. We also notice the increased magnitude of ADAS11 in AD as compared to the other clinical groups. Finally, the magnitude of change of FDG is generally milder than the atrophy rates.

The observations presented in Table 1 provide us with a glimpse into the biomarkers' trajectories characterising AD. The complexity of the dynamical changes we may infer is however limited, as the clinical stages roughly approximate a temporal scale describing the disease history, while very little insights can be obtained about the biomarkers' interactions. Within this context, our model allows the quantification of the fine-grained dynamical relationships across biomarkers at stake during the history of the disease. Investigation of intervention scenarios can be subsequently carried out by opportunely modulating the estimated dynamics parameters according to specific intervention hypothesis (e.g. amyloid lowering at a certain time).

2.2. Model overview

We provide in Figure 1 an overview of the presented method. Baseline multi-modal imaging and clinical information for a given subject are transformed into a latent variable composed of four z-scores quantifying respectively the overall severity of atrophy, glucose hypometabolism, amyloid burden, and cognitive and functional assessment. The model estimates the dynamical relationships across these z-scores to optimally describe the temporal transitions between follow-up observations. These transition rules are here mathematically defined by the parameters of a system of ODEs, which is estimated from the data. This dynamical system allows to compute the evolution of the z-scores over time from any baseline observation, and to predict the associated multi-modal imaging and clinical measures. The model thus enables to simulate the pathological progression of biomarkers across the entire history of the disease. Once the model is estimated, we can modify the ODEs parameters to simulate different evolution scenarios according to specific hypothesis. For example, by reducing the parameters associated with the progression rate of amyloid, we can investigate the relative change in the evolution of the other biomarkers. This setup thus provides us with a data-driven system enabling the exploration of hypothetical intervention strategies, and their effect on the pathological cascade.

In the following sections, MRI, FDG-PET, and AV45-PET images are processed in order to respectively extract regional gray matter density, glucose hypometabolism and amyloid load from a brain parcellation. The z-scores of gray matter atrophy (z^{atr}), glucose hypometabolism (z^{hmet}), and amyloid burden (z^{amy}), are computed using the measures obtained by this pre-processing step. The clinical z-score z^{cli} is derived from neuro-psychological scores: ADAS11, MMSE, FAQ, RAVLT immediate, and RAVLT forgetting. Further details about experimental setup, method formulation, and data pre-processing are given in Section 4.

Table 1: A: Baseline socio-demographic information for training cohort (311 subjects for 2188 data points, follow-up from 0 to 6 years depending on subjects). Average values, standard deviation in parenthesis. B: Baseline values (bl) and annual rates of change (% change / year) of amyloid burden (average normalized AV45 uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), glucose hypometabolism (average normalized FDG uptake in frontal cortex, anterior cingulate, precuneus and parietal cortex), hippocampus volume, medial temporal lobe volume, and ADAS11 score for the different clinical groups. Median values, interquartile range below. The volumes of the hippocampus and the medial temporal lobe are averaged across left and right hemispheres. NL: healthy individuals, MCI: individuals with mild cognitive impairment, AD: patients with Alzheimer's dementia. FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging. AV45: (18)F-florbetapir Amyloid PET imaging. SUVR: Standardized Uptake Value Ratio. MTL: Medial Temporal Lobe. ADAS11: Alzheimer's Disease Assessment Scale-cognitive subscale, 11 items.

A: Socio-demographics

	NL stable	NL converters	MCI stable	MCI converters	AD
N	46	10	106	76	73
Age (yrs)	73 (7)	80 (4)	73 (6)	72 (6)	74 (8)
Education (yrs)	17 (2)	15 (2)	16 (3)	16 (3)	16 (3)

B: Biomarkers and rates of change

	NL stable		NL converters		MCI stable		MCI converters		AD	
	bl	% change / year	bl	% change / year	bl	% change / year	bl	% change / year	bl	% change / year
Global AV45 (SUVR)	1.25 [1.16 ; 1.40]	1.7 [0.3 ; 2.6]	1.40 [1.26 ; 1.58]	-0.5 [-1.9 ; 1.1]	1.30 [1.21 ; 1.44]	1.5 [0.4 ; 2.6]	1.41 [1.29 ; 1.55]	0.1 [-1.4 ; 1.7]	1.42 [1.34 ; 1.54]	1.1 [-1.5 ; 2.5]
Global FDG (SUVR)	1.34 [1.27 ; 1.44]	-1.3 [-2.0 ; 0.8]	1.33 [1.27 ; 1.36]	-1.5 [-2.4 ; 0.9]	1.31 [1.24 ; 1.35]	-1.3 [-3.0 ; 0.0]	1.15 [1.07 ; 1.25]	-3.7 [-5.6 ; -1.6]	1.14 [1.07 ; 1.16]	-5.0 [-5.5 ; -1.3]
Hippocampus (ml)	3.9 [3.5 ; 4.1]	-1.6 [-2.9 ; -0.5]	3.5 [3.4 ; 3.6]	-0.4 [-2.7 ; -0.1]	3.4 [3.1 ; 3.8]	-2.5 [-3.7 ; -0.7]	3.3 [2.9 ; 3.5]	-3.8 [-5.3 ; -2.0]	2.9 [2.7 ; 3.3]	-5.1 [-8.0 ; -2.4]
MTL (ml)	10.4 [9.9 ; 11.1]	-0.8 [-2.0 ; 0.0]	9.7 [9.5 ; 9.8]	-2.2 [-6.5 ; -1.1]	10.1 [9.2 ; 11.0]	-1.1 [-2.2 ; 0.2]	9.5 [8.7 ; 10.7]	-3.0 [-5.3 ; -1.5]	8.5 [7.8 ; 9.8]	-5.9 [-7.9 ; -3.1]
ADAS11	5.5 [3.0 ; 7.8]	0.0 [-0.1 ; 0.4]	7.5 [6.0 ; 9.0]	0.7 [0.0 ; 1.6]	9.0 [6.0 ; 11.0]	1.1 [0.3 ; 3.2]	12.0 [9.0 ; 16.0]	5.2 [2.7 ; 10.3]	19.0 [15.0 ; 23.0]	7.8 [3.8 ; 17.8]

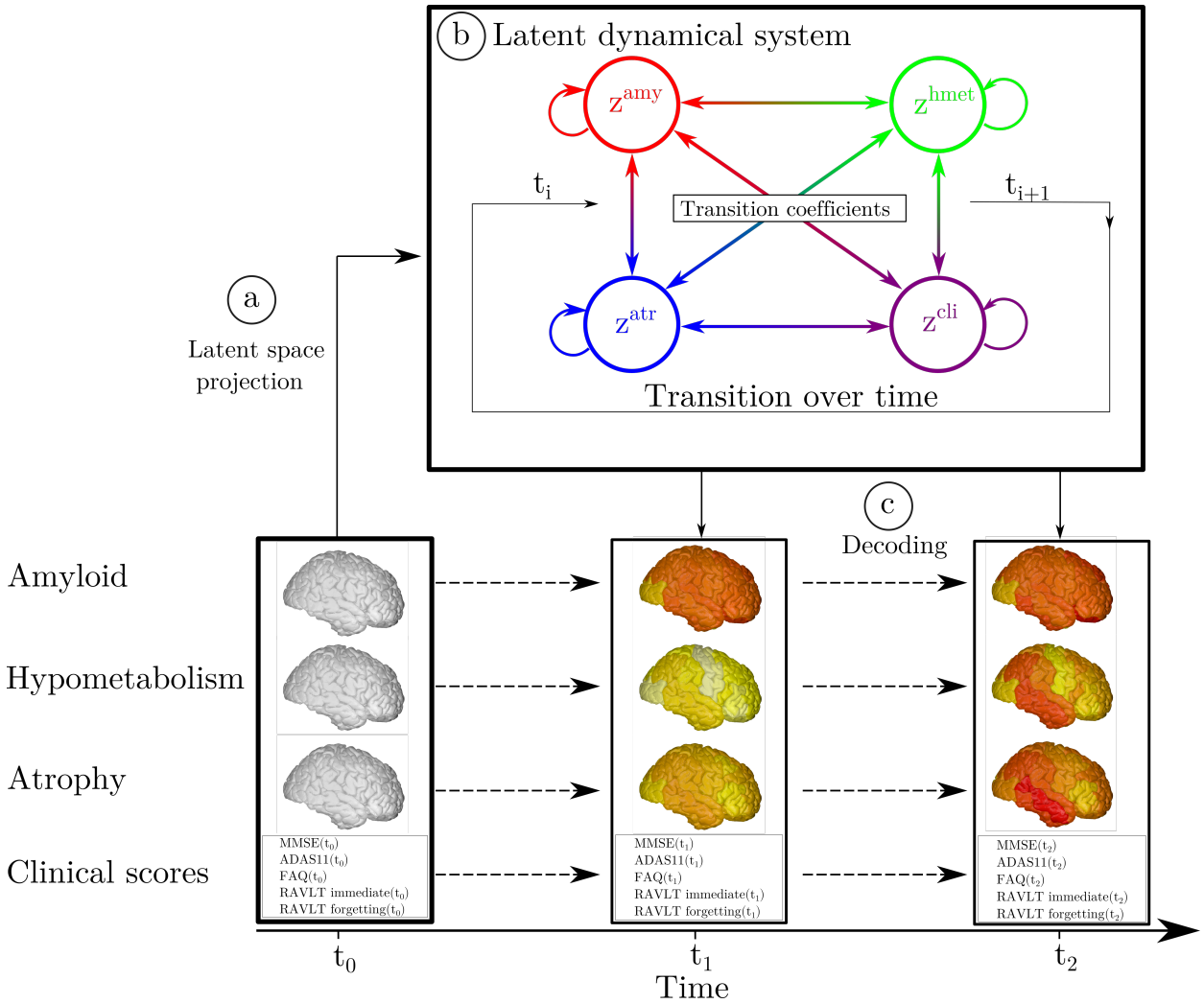

Figure 1: Overview of the method. a) High-dimensional multi-modal measures are projected into a 4-dimensional latent space. Each data modality is transformed in a corresponding z-score $z^{amy}, z^{hmet}, z^{atr}, z^{cli}$. b) The dynamical system describing the relationships between the z-scores allows to compute their transition across the evolution of the disease. c) Given the latent space and the estimated dynamics, the follow-up measurements can be reconstructed to match the observed data.

2.3. Progression model and latent relationships

We show in Figure 2 the dynamical relationships across the different z-scores estimated by the model, where direction and intensity of the arrows quantify the estimated increase of one variable with respect to the other. Being the scores adimensional, they have been conveniently rescaled to the range [0,1] indicating increasing pathological levels. These relationships extend the summary statistics reported in Table 1 to a much finer temporal scale and wider range of possible biomarkers' values. We observe in Figure 2A, 2B and 2C that large values of the amyloid score z^{amy} trigger the increase of the remaining ones: z^{hmet}, z^{atr}, and z^{cli}. Figure 2D shows that large increase of the atrophy score z^{atr} is associated to higher hypometabolism indicated by large values of z^{hmet}. Moreover, we note that high z^{hmet} values also contribute to an increase of z^{cli} (Figure 2E). Finally, Figure 2F shows that high atrophy values lead to an increase mostly along the clinical dimension z^{cli}. This chain of relationships is in agreement with the cascade hypothesis of AD [2, 3].

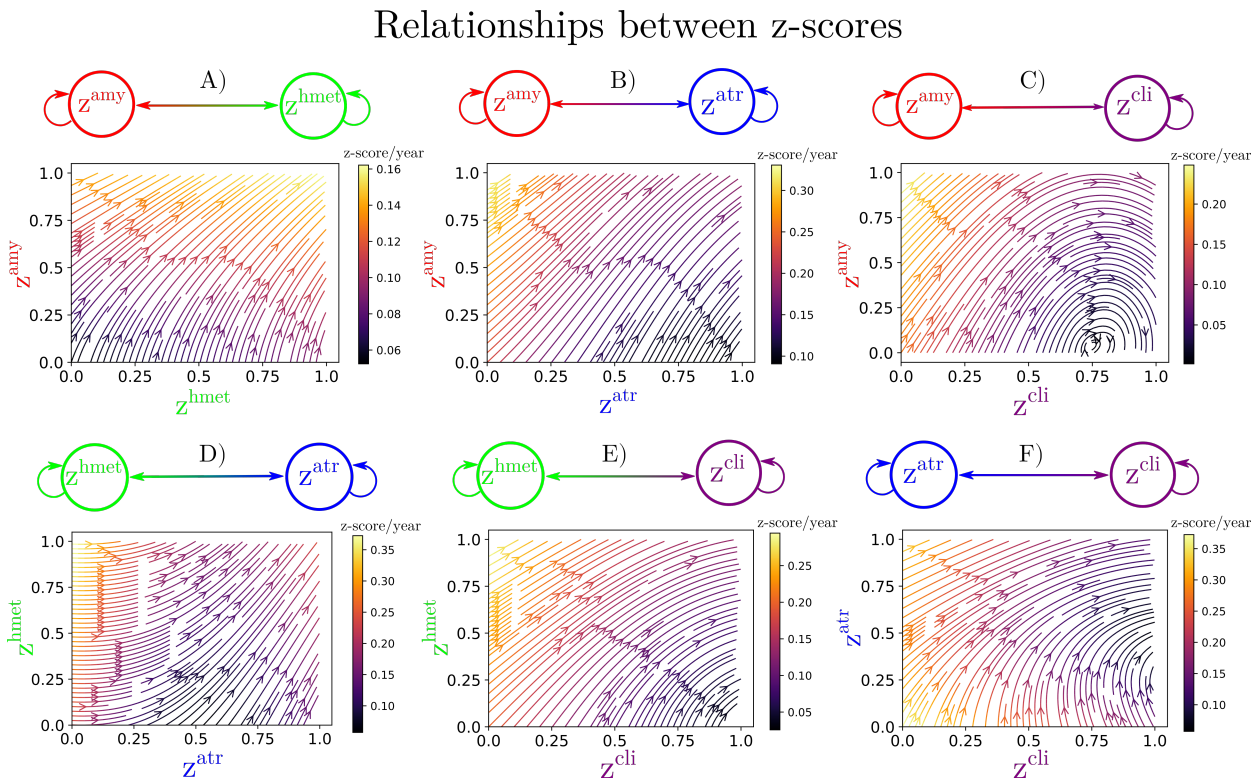

Figure 2: Estimated dynamical relationships across the different z-scores (A to F). Given the values of two z-scores, the arrow at the corresponding coordinates indicates how one score evolves with respect to the other. The intensity of the arrow gives the strength of the relationship between the two scores.

Relying on the dynamical relationships shown in Figure 2, starting from any initial set of biomarkers values we can estimate the relative trajectories over time. Figure 3 (left), shows the evolution obtained by extrapolating backward and forward in time the trajectory associated to the z-scores of the AD group. The x-axis represents the years from conversion to AD, where the instant t=0 corresponds to the average time of diagnosis estimated for the group of MCI progressing to dementia. As observed in Figure 2 and Table 1, the amyloid score z^{amy} increases and saturates first, followed by z^{hmet} and z^{atr} scores whose progression slows down when reaching clinical conversion, while the clinical score exhibits strong acceleration in the latest progression stages. Figure 3 (right) shows the group-wise distribution of the disease severity estimated for each subject relatively to the modelled long-term latent trajectories (Section 4.7). The group-wise difference of disease severity across groups is statistically significant and increases when going from healthy to pathological stages (Wilcoxon-Mann-Whitney test $p < 0.001$ for each comparisons). The reliability of the estimation of disease severity was further assessed through testing on an independent cohort, and by comparison with a

previously proposed disease progression modeling method from the state-of-the-art [25]. The results are provided in section 1 of the Supplementary Material and show positive generalization results as well as a favourable comparison with the benchmark method.

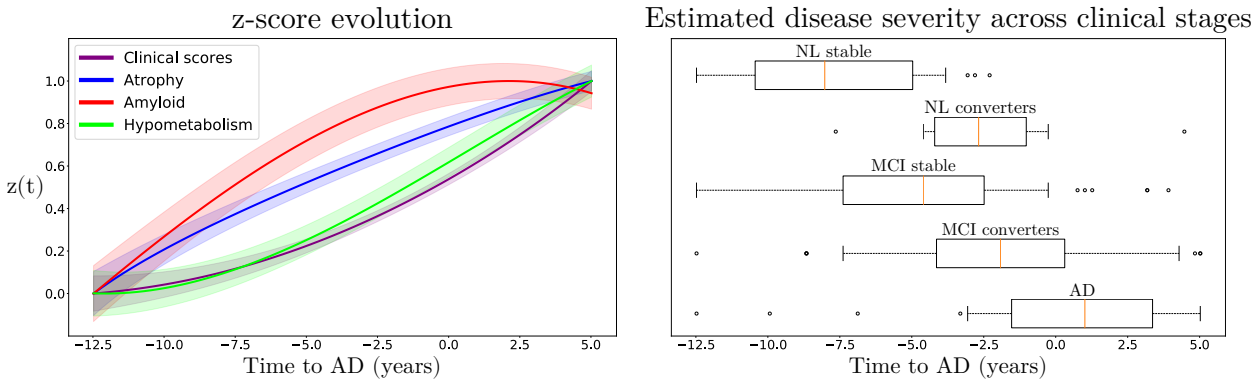

Figure 3: Left: Estimated long-term latent dynamics (time is relative to conversion to Alzheimer's dementia). Shadowed areas represent the standard deviation of the average trajectory. Right: Distribution of the estimated disease severity across clinical stages, relatively to the long-term dynamics on the left. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer's dementia.

From the z-score trajectories of Figure 3 (left) we predict the progression of imaging and clinical measures shown in Figure 4. We observe that amyloid load globally increases and saturates early, compatibly with the positive amyloid condition of the study cohort. Glucose hypometabolism and gray matter atrophy increase are delayed with respect to amyloid, and tend to map prevalently temporal and parietal regions. Finally, the clinical measures exhibit a non-linear pattern of change, accelerating during the latest progression stages. These dynamics are compatible with the summary measures on the raw data reported in Table 1.

2.4. Simulating clinical intervention

This experimental section is based on two intervention scenarios: a first one in which amyloid is lowered by 100%, and a second one in which it is reduced by 50% with respect to the estimated natural progression. In Figure 5 we show the latent z-scores evolution resulting from either 100% or 50% amyloid lowering performed at the time $t = -12.5$ years. According to these scenarios, intervention results in a sensitive reduction of the pathological progression for atrophy, hypometabolism and clinical scores, albeit with a stronger effect in case of total blockage.

We further estimated the resulting clinical endpoints associated with the two amyloid lowering scenarios, at increasing time points and for different sample sizes. Clinical endpoints consisted in the simulated ADAS11, MMSE, FAQ, RAVLT immediate, and RAVLT forgetting scores at the reference conversion time (t=0). The case placebo indicates the scenario where clinical values were computed at conversion time from the estimated natural progression shown in Figure 3. Figure 6 shows the change in statistical power depending on intervention time and sample sizes. For large sample sizes (1000 subjects per arm) a power greater than 0.8 can be obtained around 5 years before conversion, depending on the outcome score, where in general we observe that RAVLT forgetting exhibits a higher power than the other scores. When sample size is lower than 100 subjects per arm, a power greater than 0.8 is reached if intervention is performed at the latest 8 years before conversion, with a mild variability depending on the considered clinical score. We notice that in the case of a 50% amyloid lowering, in order to reach the same power intervention needs to be consistently performed earlier compared to the scenario of 100% amyloid lowering for the same sample size and clinical score. For instance, if we consider ADAS11 with a sample size of 100 subjects per arm, a power of 0.8 is obtained for a 100% amyloid lowering intervention performed 8 years before conversion, while in case of a 50% amyloid lowering the equivalent effect would be obtained by intervening 10.5 years before conversion.

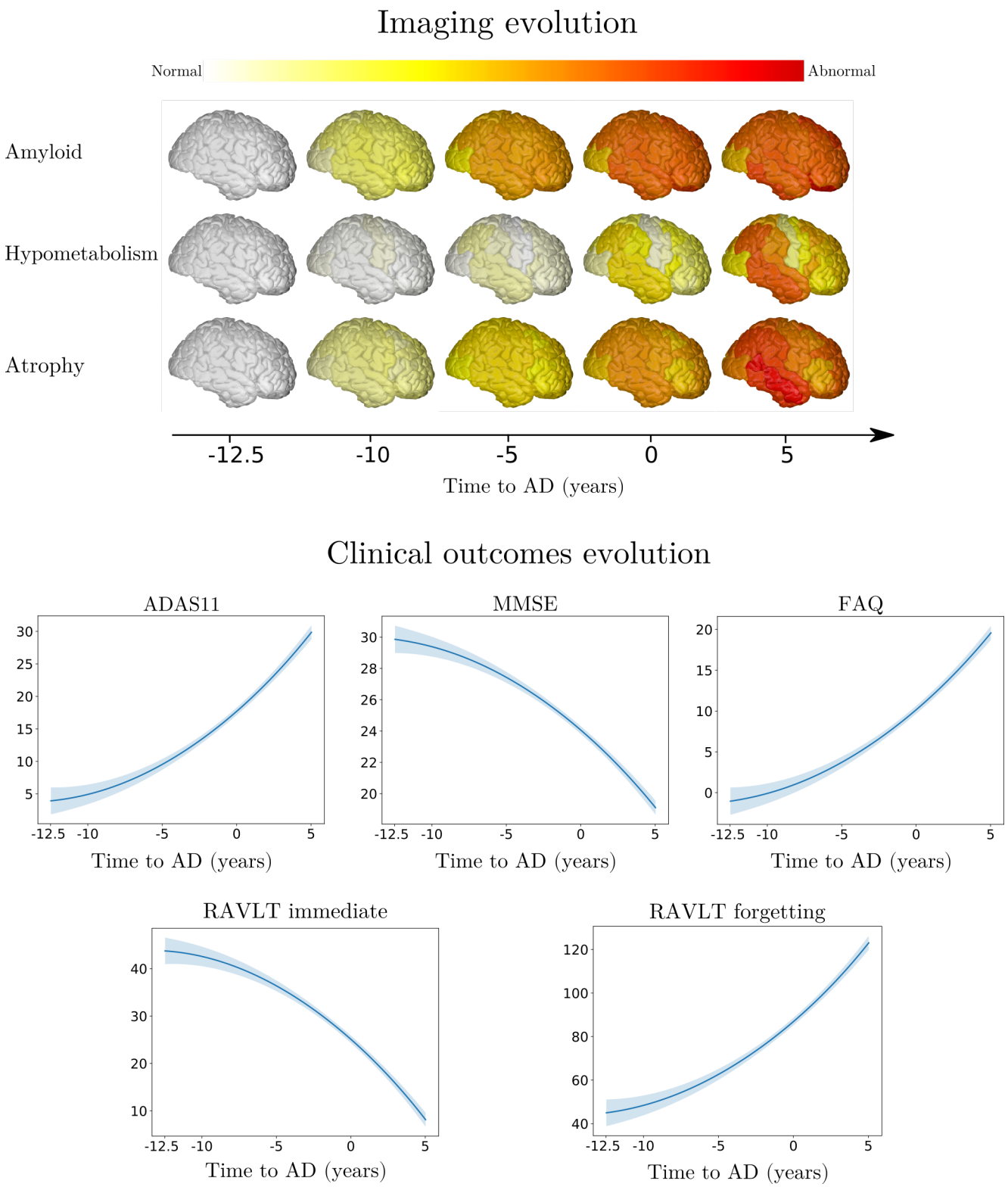

Figure 4: Modelled long-term evolution of cortical measurements for the different types of imaging markers, and clinical scores. Shadowed areas represent the standard deviation of the average trajectory. Brain images were generated using the software provided in [38].

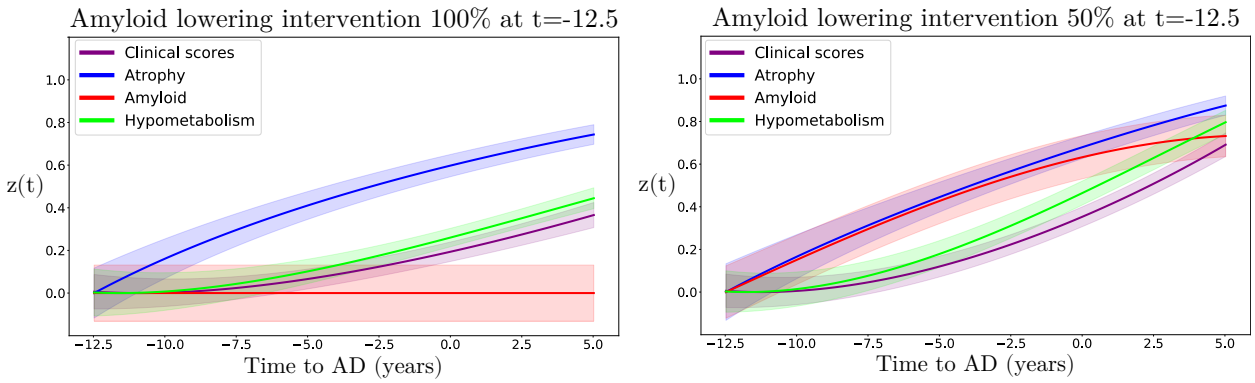

Figure 5: Hypothetical scenarios of irreversible amyloid lowering interventions at t=-12.5 years from Alzheimer's dementia diagnosis, with a rate of 100 % (left) or 50 % (right). Shadowed areas represent the standard deviation of the average trajectory.

We provide in Table 2 the estimated improvement for each clinical score at conversion with a sample size of 100 subjects per arm for both 100% and 50% amyloid lowering depending on the intervention time. We observe that for the same intervention time, 100% amyloid lowering always results in a larger improvement of clinical endpoints compared to 50% amyloid lowering. We also note that in the case of 100% lowering, clinical endpoints obtained for intervention at t=-10 years correspond to typical cutoff values for inclusion into AD trials (ADAS11= 13.4 ± 6.2, MMSE= 25.8 ± 2.5, see Supplementary Table 2) [39, 40].

Table 2: Estimated mean (standard deviation) improvement of clinical outcomes at predicted conversion time for the normal progression case by year of simulated intervention (100% and 50% amyloid lowering interventions). Results in bold indicate a statistically significant difference between placebo and treated scenarios (p<0.01, two-sided t-test, 100 cases per arm). AD: Alzheimer's dementia, ADAS11: Alzheimer's Disease Assessment Scale, MMSE: Mini-Mental State Examination, FAQ: Functional Assessment Questionnaire, RAVLT: Rey Auditory Verbal Learning Test.

Amyloid lowering intervention 100%
Point improvement per intervention time

Score \ Years to AD	−12.5	−10	−7.5	−5	−4	−3	−2	−1
ADAS11	**7.0 (4.8)**	**4.3 (2.8)**	**2.0 (1.4)**	0.7 (0.5)	0.4 (0.3)	0.2 (0.2)	0.1 (0.1)	0.01 (0.01)
MMSE	**2.8 (1.9)**	**1.8 (1.1)**	**0.9 (0.6)**	0.3 (0.2)	0.2 (0.1)	0.1 (0.06)	0.02 (0.02)	0.0 (0.0)
FAQ	**5.5 (3.8)**	**3.4 (2.3)**	**1.7 (1.1)**	0.6 (0.5)	0.4 (0.3)	0.2 (0.2)	0.1 (0.1)	0.01 (0.01)
RAVLT immediate	**10.0 (7.0)**	**6.2 (4.1)**	**3.0 (2.0)**	1.2 (0.8)	0.7 (0.5)	0.3 (0.3)	0.1 (0.1)	0.03 (0.02)
RAVLT forgetting (%)	**23.6 (16.2)**	**15.0 (10.0)**	**7.7 (5.1)**	3.0 (2.2)	2.0 (1.6)	1.0 (1.0)	0.4 (0.4)	0.2 (0.1)

Amyloid lowering intervention 50%
Point improvement per intervention time

Score \ Years to AD	−12.5	−10	−7.5	−5	−4	−3	−2	−1
ADAS11	**3.5 (2.4)**	**2.1 (1.4)**	1.0 (0.7)	0.4 (0.3)	0.2 (0.2)	0.1 (0.1)	0.04 (0.03)	0.0 (0.0)
MMSE	**1.4 (1.0)**	**0.9 (0.6)**	0.4 (0.3)	0.1 (0.1)	0.1(0.1)	0.03 (0.02)	0.0 (0.0)	0.0 (0.0)
FAQ	**2.8 (1.9)**	**1.7 (1.1)**	0.8 (0.5)	0.3 (0.2)	0.2 (0.1)	0.1 (0.1)	0.03 (0.03)	0.0 (0.0)
RAVLT immediate	**5.0 (3.5)**	**3.1 (2.0)**	1.5 (1.0)	0.6 (0.4)	0.3 (0.3)	0.2 (0.1)	0.1 (0.1)	0.01 (0.01)
RAVLT forgetting (%)	**11.8 (8.1)**	**7.4 (4.9)**	3.7 (2.5)	1.5 (1.0)	0.9 (0.7)	0.5 (0.5)	0.3 (0.2)	0.1 (0.1)

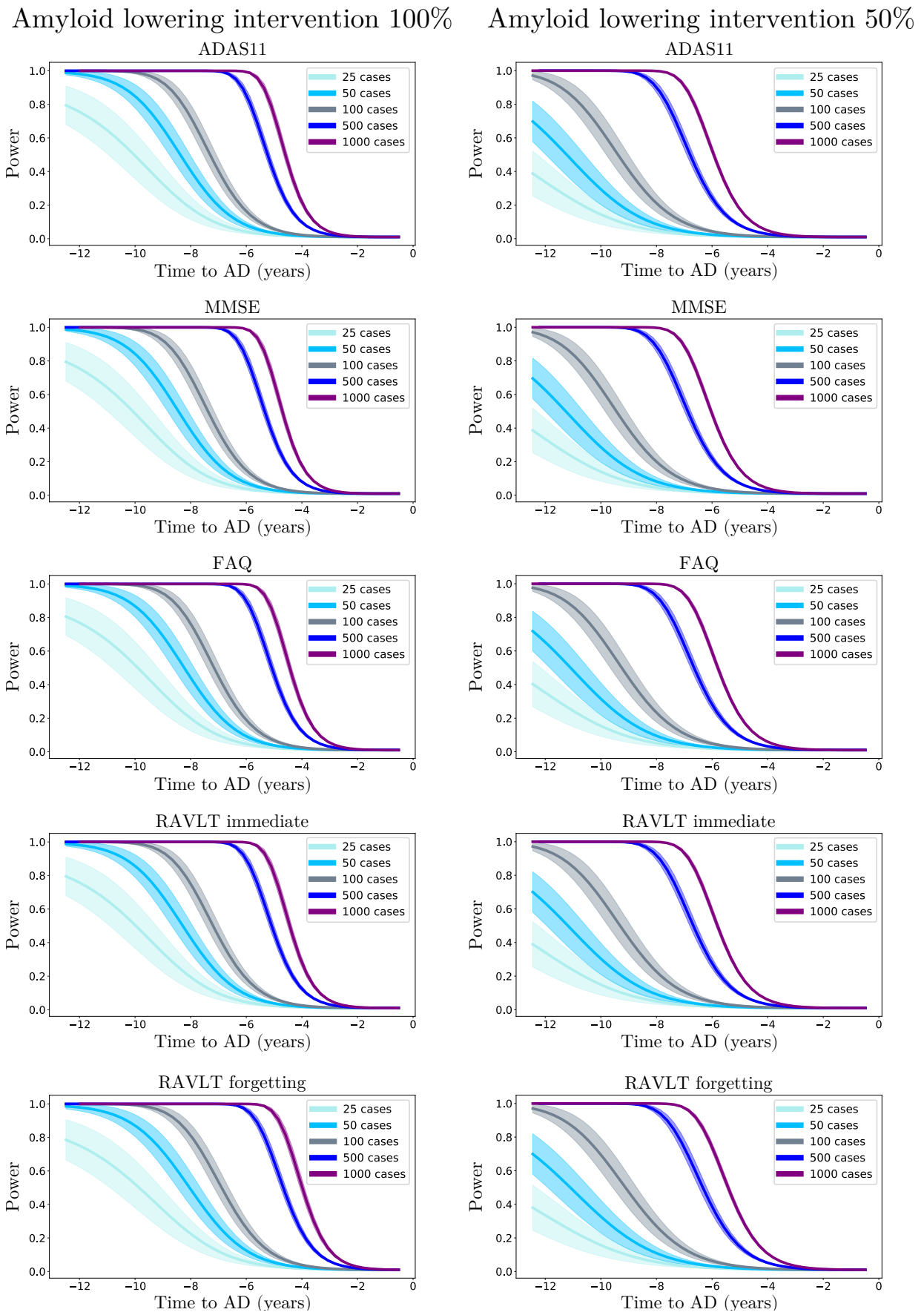

Figure 6: Statistical power of the Student t-test comparing the estimated clinical outcomes at conversion time between placebo and treated scenarios, according to the year of simulated intervention (100% and 50% amyloid lowering) and sample size.

3. Discussion

We presented a framework to jointly model the progression of multi-modal imaging and clinical data, based on the estimation of latent biomarkers' relationships governing AD progression. The model is designed to simulate intervention scenarios in clinical trials, and in this study we focused on assessing the effect of anti-amyloid drugs on biomarkers' evolution, by quantifying the effect of intervention time and drug efficacy on clinical outcomes. Our results underline the critical importance of intervention time, which should be performed sensibly early during the pathological history to effectively appreciate the effectiveness of disease modifiers.

The results obtained with our model are compatible with findings reported in recent clinical studies [11, 12, 13]. For example, if we consider 500 patients per arm and perform a 100% amyloid lowering intervention for 2 years to reproduce the conditions of the recent trial of Verubecestat [12], the average improvement of MMSE predicted by our model is of 0.02, falling in the 95% confidence interval measured during that study ([-0.5 ; 0.8]). While recent anti-amyloid trials such as [11, 12, 13] included between 500 and 1000 mild AD subjects per arm and were conducted over a period of two years at most, our analysis suggests that clinical trials performed with less than 1000 subjects with mild AD may be consistently under-powered. Indeed, we see in Figure 6 that with a sample size of 1000 subjects per arm and a total blockage of amyloid production, a power of 0.8 can be obtained only if intervention is performed at least 5 years before conversion.

These results allow to quantify the crucial role of intervention time, and provide an experimental justification for testing amyloid modifying drugs in the pre-clinical stage [15, 41]. This is for example illustrated in Table 2, in which we notice that clinical endpoints are close to placebo even when the simulated intervention takes place up to 5 years before conversion, while stronger cognitive and functional changes happen when amyloid is lowered by 100% or 50% at least 10 years before conversion. These findings may be explained by considering that amyloid accumulates over more than a decade, and that when amyloid clearance occurs the pathological cascade is already entrenched [42]. Our results are thus supporting the need to identify subjects at the pre-clinical stage, that is to say still cognitively normal, which is a challenging task. Currently, one of the main criteria to enroll subjects into clinical trials is the presence of amyloid in the brain, and blood-based markers are considered as potential candidates for identifying patients at risk for AD [43]. Moreover, recent works such as [44, 45] have proposed more complex entry criteria to constitute cohorts based on multi-modal measurements. Within this context, our model could also be used as an enrichment tool by quantifying the disease severity based on multi-modal data as shown in Figure 3. Similarly, the method could be applied to predict the evolution of single patient given its current available measurements.

An additional critical aspect of anti-amyloid trials is the effect of dose exposure on the production of amyloid [16]. Currently, β-site amyloid precursor protein cleaving enzyme (BACE) inhibitors allow to suppress amyloid production from 50% to 90%. In this study we showed that lowering amyloid by 50% consistently decreases the treatment effect compared to a 100% lowering at the same time. For instance, if we consider a sample size of 1000 subjects per arm in the case of a 50% amyloid lowering intervention, an 80% power can be reached only 6.5 years before conversion instead of 5 years for a 100% amyloid lowering intervention. This ability of our model to control the rate of amyloid progression is fundamental in order to provide realistic simulations of anti-amyloid trials.

In Figure 2 we showed that amyloid triggers the pathological cascade affecting the other markers, thus confirming its dominating role on disease progression. Assuming that the data used to estimate the model is sufficient to completely depict the history of the pathology, our model can be interpreted from a causal perspective. However, we cannot exclude the existence of other mechanisms driving amyloid accumulation, which our model cannot infer from the existing data. Therefore, our findings should be considered with care, while the integration of additional biomarkers of interest will be necessary to account for multiple drivers of the disease. It is worth noting that recent works ventured the idea to combine drugs targeting multiple mechanisms at the same time [46]. For instance, pathologists have shown tau deposition in brainstem nuclei in adolescents and children [47], and clinicians are currently investigating the pathological effect of early tau spreading on AD progression [48], raising crucial questions about its relationship with amyloid accumulation, and the impact on cognitive impairment [49]. Our model would allow to address these questions by including measures derived from Tau-PET images, and simulating scenarios of production blockage of

both proteins at different rates or intervention time.

Lately, disappointing results of clinical studies led to hypothesize specific treatments targeting AD sub-populations based on their genotype [50]. While in our work we describe a global progression of AD, in the future we will account for sub-trajectories due to genetic factors, such as the presence of $\epsilon 4$ allele of apolipoprotein (APOE4), which is a major risk for developing AD influencing both disease onset and progression [51]. This could be done by estimating dynamical systems specific to the genetic condition of each patient. Simulating the dynamical relationships specific to genetic factors would allow to evaluate the effect of APOE4 on intervention time or drug dosage. In addition, there exist numerous non-genetic aggravating factors that may also affect disease evolution, such as diabetes, obesity or smoking. Extending our model to account for panels of risk factors would ultimately allow to test in silico personalized intervention strategies. Moreover, a key aspect of clinical trials is their economic cost. Our model could be extended to help designing clinical trials by optimizing intervention with respect to the available funding. Given a budget, we could simulate scenarios based on different sample size, and trials duration, while estimating the expected cognitive outcome.

Results presented in this work are based on a model estimated by relying solely on a subset of the ADNI cohort, and therefore they may not be fully representative of the general AD progression. Indeed, subjects included in this cohort were either amyloid-positive at baseline, or became amyloid-positive during their follow-up visits (see Section 2.1). They may therefore provide a limited representation of the pathological temporal window captured by the model. Applying the model on a cohort containing amyloid-negative subjects may provide additional insights on the overall disease history. However, this is a challenging task as it would require to identify sub-trajectories dissociated from normal ageing [52, 53]. In addition to this specific characteristic of the cohort, there exists additional biases impacting the model estimation. For instance, the fact that gray matter atrophy becomes abnormal before glucose metabolism in Figure 4 can be explained by the generally high atrophy rate of change in some key regions in normal elders, such as in the hippocampus, compared to the rate of change of FDG (see Table 1). We note that this stronger change of atrophy with respect to glucose hypometabolism can already be appreciated in the clinically healthy group. The existence of such biases can also be observed in Figure 5, in which we notice that atrophy is less affected by intervention, implying that its evolution is here importantly decorrelated from amyloid burden.

4. Methods

4.1. Data acquisition and preprocessing

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. For up-to-date information, see www.adni-info.org.

We considered four types of biomarkers, related to clinical scores, gray matter atrophy, amyloid load and glucose hypometabolism, and respectively denoted by *cli*, *atr*, *amy* and *hmet*. MRI images were processed following the longitudinal pipeline of Freesurfer [54], to obtain gray matter volumes in a standard anatomical space. AV45-PET and FDG-PET images were aligned to the closest MRI in time, and normalized to the cerebellum uptake. Regional gray matter density, amyloid load and glucose hypometabolism were extracted from the Desikan-Killiany parcellation [55]. We discarded white-matter, ventricular, and cerebellar regions, thus obtaining 82 regions that were averaged across hemispheres. Therefore, for a given subject, $\mathbf{x}^{atr}, \mathbf{x}^{amy}$ and $\mathbf{x}^{hmet}$ are respectively 41-dimensional vectors. The variable $\mathbf{x}^{cli}$ is composed of the neuro-psychological scores ADAS11, MMSE, RAVLT immediate, RAVLT forgetting and FAQ. The total number of measures is of 2188 longitudinal data points. We note that the model requires all the measures to be available at baseline in order to obtain a latent representation, but is able to handle missing data in the follow-up. Further details on the cohort are given in Section 2.1.

4.2. Data modelling

We consider observations $\mathbf{X}_i(t) = [\mathbf{x}_i^1(t), \mathbf{x}_i^2(t), ..., \mathbf{x}_i^M(t)]^T$, which correspond to multivariate measures derived from M different modalities (e.g clinical scores, MRI, AV45, or FDG measures) at time t for subject i. Each vector $\mathbf{x}_i^m(t)$ has

dimension D_m. We postulate the following generative model, in which the modalities are assumed to be independently generated by a common latent representation of the data $\mathbf{z}_i(t)$:

$$\begin{aligned} p(\mathbf{X}_i(t)|\mathbf{z}_i(t), \boldsymbol{\sigma^2}, \boldsymbol{\psi}) &= \prod_m p(\mathbf{x}_i^m(t)|\mathbf{z}_i(t), \sigma_m^2, \psi_m) \\ &= \prod_m \mathcal{N}(\mu_m(\mathbf{z}_i(t), \psi_m), \sigma_m^2), \\ \mathbf{z_i}(t) &= \Lambda(\mathbf{z}_i(t_0), t), \\ \mathbf{z}_i(t_0) &\sim p(\mathbf{z}_i(t_0)), \end{aligned} \tag{1}$$

where σ_m^2 is measurement noise, while ψ_m are the parameters of the function μ_m which maps the latent state to the data space for the modality m. For simplicity of notation we denote $\mathbf{z}_i(t)$ by $\mathbf{z}(t)$. We assume that each coordinate of $\mathbf{z}$ is associated to a specific modality m, leading to an M-dimensional latent space. The Λ operator which gives the value of the latent representation at a given time t, is defined by the solution of the following system of ODEs:

$$\frac{dz^m(t)}{dt} = k_m z^m(t)(1 - z^m(t)) + \sum_{j \neq m} \alpha_{m,j} z^j(t), \quad m{=}1,...,M. \tag{2}$$

For each coordinate, the first term of the equation enforces a sigmoidal evolution with a progression rate k_m, while the second term accounts for the relationship between modalities m and j through the parameters $\alpha_{m,j}$. This system can be rewritten as:

$$\begin{aligned} &\frac{d\mathbf{z}(t)}{dt} = \mathbf{W}\mathbf{z}(t) - \mathbf{V}\mathbf{z}^2(t) = g(\mathbf{z}(t), \theta_{ODE}) \text{ where,} \\ &\left(\mathbf{W}_{i,j}\right) = \begin{cases} k_i \text{ if i=j,} \\ \alpha_{i,j} \text{ otherwise;} \end{cases} \text{ and } \left(\mathbf{V}_{i,j}\right) = \begin{cases} k_i \text{ if i=j} \\ 0 \text{ otherwise,} \end{cases} \end{aligned} \tag{3}$$

θ_{ODE} denotes the parameters of the system of ODEs, which correspond to the entries of the matrices $\mathbf{W}$ and $\mathbf{V}$. According to Equation 3, for each initial condition $\mathbf{z}(0)$, the latent state at time t can be computed through integration, $\mathbf{z}(t) = \mathbf{z}(0) + \int_0^t g(\mathbf{z}(x), \theta_{ODE})dx$.

4.3. Variational inference

We rewrite $p(\mathbf{X}_i(t)|\mathbf{z}_i(t), \boldsymbol{\sigma^2}, \boldsymbol{\psi})$ as $p(\mathbf{X}_i(t)|\mathbf{z}_i(t_0), \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi})$. Assuming independence between subjects, the marginal log-likelihood writes as:

$$\begin{aligned} \mathcal{L} &= \sum_i^N \log\left[p(\mathbf{X}_i(t)|\theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi})\right] \\ &= \sum_i^N \log\left[\int p(\mathbf{X}_i(t)|\mathbf{z}_i(t_0), \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) p(\mathbf{z}_i(t_0)) d\mathbf{z}_i(t_0)\right]. \end{aligned} \tag{4}$$

For ease of notation, we drop the i index, and dependence on t and t_0 is made implicit. Within a Bayesian framework, we wish to maximize $\mathcal{L}$ in order to obtain a posterior distribution for the latent variable $\mathbf{z}$. Since derivation of this quantity is generally not tractable, we resort to stochastic variational inference to tackle the optimization problem. We assume a $\mathcal{N}(\mathbf{0}, \mathbf{I})$ prior for $p(\mathbf{z})$, and introduce an approximate posterior distribution $q(\mathbf{z}|\mathbf{X})$ [56], in order to derive a lower-bound $\mathcal{E}$ for the marginal log-likelihood:

$$
\begin{aligned}
\log p(\mathbf{X}|\theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{X})} \Big[\log p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) \Big] \\
- \mathcal{D}\Big[q(\mathbf{z}|\mathbf{X}) | p(\mathbf{z}) \Big] \\
= \mathcal{E},
\end{aligned} \tag{5}
$$

where $\mathcal{D}$ refers to the Kullback-Leibler (KL) divergence. We propose to factorize the distribution $q(\mathbf{z}|\mathbf{X})$ across modalities such that, $q(\mathbf{z}|\mathbf{X}) = \prod_m q(z^m|\mathbf{x}^m)$, where $q(z^m|\mathbf{x}^m) = \mathcal{N}(f(\mathbf{x}^m, \phi_m^1), h(\mathbf{x}^m, \phi_m^2))$, is a variational Gaussian approximation with moments parameterized by the functions f and h. This modality-wise encoding of the data enables to interpret each coordinate of $\mathbf{z}$ as a compressed representation of the corresponding modality. Moreover, the lower-bound simplifies as:

$$
\mathcal{E} = \sum_m \mathbb{E}_{q(\mathbf{z}|\mathbf{X})} \Big[\log p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m) \Big] - \mathcal{D}\Big[q(z^m|\mathbf{x}^m) | p(z^m) \Big]. \tag{6}
$$

Details about the ELBO derivation and the computation of the KL divergence are given in sections 3 and 4 of the Supplementary Material. A graphical model of the method is also provided in Supplementary Figure 3, while Supplementary Algorithm 1 details the steps to compute the ELBO.

4.4. Model optimization

Using the reparameterization trick [57], we can efficiently sample from the posterior distribution $q(\mathbf{z}(t_0)|\mathbf{X}(t_0))$ to approximate the expectation terms. Moreover, thanks to our choices of priors and approximations the KL terms can be computed in closed-form. In practice, we sample from $q(\mathbf{z}(t_0)|\mathbf{X}(t_0))$ to obtain a latent representation $\mathbf{z}(t_0)$ at baseline, while the follow-up points are estimated by decoding the latent time-series obtained through the integration of the ODEs of Eq 3. The model is trained by computing the total ELBO for all the subjects at all the available time points. The parameters $\boldsymbol{\psi}, \boldsymbol{\phi}^1, \boldsymbol{\phi}^2, \theta_{ODE}, \boldsymbol{\sigma}$ are optimized using gradient descent, which requires to backpropagate through the integration operation.

In order to enable backpropagation through the ODEs integration we need to numerically solve the differential equation using only operations that can be differentiated. In this work, we used the Midpoint method which follows a second order Runge-Kutta scheme. The method consists in evaluating the derivative of the solution at $(t_{i+1} + t_i)/2$, which is the midpoint between t_i at which the correct $\mathbf{z}(t)$ is evaluated, and the following t_{i+1}:

$$
\begin{aligned}
\int_{t_i}^{t_{i+1}} g(\mathbf{z}(x))dx &\approx h \cdot g\Big(\mathbf{z}(\frac{t_i + t_{i+1}}{2})\Big) \\
&\approx h \cdot g\Big(\mathbf{z}(t_i) + \frac{h}{2} g(\mathbf{z}(t_i))\Big), \quad h = t_{i+1} - t_i.
\end{aligned} \tag{7}
$$

Therefore, solving the system of Equation 3 on the interval $[t_0, ..., t]$ only requires operations that can be differentiated, allowing to compute the derivatives of the ELBO with respect to all the parameters, and to optimize them by gradient descent. Moreover, in order to control the variability of the estimated latent trajectory $\mathbf{z}(t)$ due to the error propagation during integration, we initialized the weights of $\boldsymbol{\phi}^1$ and $\boldsymbol{\phi}^2$ such that the approximate posterior of the latent representation for each modality m at baseline was following a $\mathcal{N}(0, 0.01)$ distribution. Finally, we also tested other ODE solvers such as Runge-Kutta 4, which gave similar results than the Midpoint method with a slower execution time due its more expensive approximation scheme.

Concerning the implementation, we trained the model using the ADAM optimizer [58] with a learning rate of 0.01. The functions f, h and μ_m were parameterized as linear transformations. The model was implemented in Pytorch [59], and we used the *torchdiffeq* package developed in [60] to backpropagate through the ODE solver.

4.5. Simulating the long-term progression of AD

To simulate the long-term progression of AD we first project the AD cohort in the latent space via the encoding functions. We can subsequently follow the trajectories of these subjects backward and forward in time, in order to estimate the associated trajectory from the healthy to their respective pathological condition. In practice, a Gaussian Mixture Model is used to fit the empirical distribution of the AD subjects' latent projection. The number of components and covariance type of the GMM is selected by relying on the Akaike information criterion [61]. The fitted GMM allows us to sample pathological latent representations $\mathbf{z}_i(t_0)$, that can be integrated forward and backward in time thanks to the estimated set of latent ODEs, to finally obtain a collection of latent trajectories $\mathbf{Z}(t) = [\mathbf{z}_1(t), ..., \mathbf{z}_N(t)]$ summarising the distribution of the long-term AD evolution.

4.6. Simulating intervention

In this section we assume that we computed the average latent progression of the disease $\mathbf{z}(t)$. Thanks to the modality-wise encoding of Section 4.3 each coordinate of the latent representation can be interpreted as representing a single data modality. Therefore, we propose to simulate the effect of an hypothetical intervention on the disease progression, by modulating the vector $\frac{d\mathbf{z}(t)}{dt}$ after each integration step such that:

$$\left(\frac{d\mathbf{z}(t)}{dt}\right)^* = \mathbf{\Gamma}\frac{d\mathbf{z}(t)}{dt} \text{ where, } \mathbf{\Gamma} = \begin{pmatrix} \gamma_1 & & \\ & \ddots & \\ & & \gamma_m \end{pmatrix}. \tag{8}$$

The values γ_m are fixed between 0 and 1, allowing to control the influence of the corresponding modalities on the system evolution, and to create hypothetical scenarios of evolution. For example, for a 100% (resp. 50%) amyloid lowering intervention we set $\gamma_{amy} = 0$ (resp. $\gamma_{amy} = 0.5$).

4.7. Evaluating disease severity

Given an evolution $\mathbf{z}(t)$ describing the disease progression in the latent space, we propose to consider this trajectory as a reference and to use it in order to quantify the individual disease severity of a subject $\mathbf{X}$. This is done by estimating a time-shift τ defined as:

$$\begin{aligned} \tau &= \arg\min_t \|f(\mathbf{X}, \boldsymbol{\phi}^1) - \mathbf{z}(t)\|_1 \\ &= \sum_m |f(\mathbf{x}^m, \phi^1) - z^m(t)|. \end{aligned} \tag{9}$$

This time-shift allows to quantify the pathological stage of a subject with respect to the disease progression along the reference trajectory $\mathbf{z}(t)$. Moreover, the time-shift can still be estimated even in the case of missing data modalities, by only encoding the available measures of the observed subject.

Data availability

The data used in this study are available from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu).

References

[1] M. Prince, A. Wimo, M. Guerchet, G. Ali, Y.-T. Wu, M. Prina, World Alzheimer Report 2015 - The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends, Alzheimer's Disease International, 2015.

[2] C. R. Jack, D. M. Holtzman, Biomarker modeling of Alzheimer's disease, Neuron 80 (2013) 1347–1358.

[3] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, J. Q. Trojanowski, Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers, Lancet Neurol 12 (2013) 207–216.

[4] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoeke, S. L. Macaulay, R. Martins, P. Maruff, D. Ames, C. C. Rowe, C. L. Masters, Amyloid Î² deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study, Lancet Neurol 12 (2013) 357–367.

[5] M. P. Murphy, H. LeVine, Alzheimer's disease and the amyloid-beta peptide, J. Alzheimers Dis. 19 (2010) 311–323.

[6] A. Delacourte, J. P. David, N. Sergeant, L. Buée, A. Wattez, P. Vermersch, F. Ghozali, C. Fallet-Bianco, F. Pasquier, F. Lebert, H. Petit, C. Di Menza, The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer's disease, Neurology 52 (1999) 1158–1165.

[7] H. Braak, E. Braak, Neuropathological stageing of Alzheimer-related changes, Acta Neuropathol. 82 (1991) 239–259.

[8] R. J. Bateman, C. Xiong, T. L. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, D. S. Marcus, N. J. Cairns, X. Xie, T. M. Blazey, D. M. Holtzman, A. Santacruz, V. Buckles, A. Oliver, K. Moulder, P. S. Aisen, B. Ghetti, W. E. Klunk, E. McDade, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, P. R. Schofield, R. A. Sperling, S. Salloway, J. C. Morris, Clinical and biomarker changes in dominantly inherited alzheimer's disease, New England Journal of Medicine 367 (2012) 795–804. PMID: 22784036.

[9] J. Cummings, G. Lee, A. Ritter, M. Sabbagh, K. Zhong, Alzheimer's disease drug development pipeline: 2019, Alzheimers Dement (N Y) 5 (2019) 272–293.

[10] R. Howard, K. Y. Liu, Questions EMERGE as Biogen claims aducanumab turnaround, Nat Rev Neurol 16 (2020) 63–64.

[11] L. S. Honig, B. Vellas, M. Woodward, M. Boada, R. Bullock, M. Borrie, K. Hager, N. Andreasen, E. Scarpini, H. Liu-Seifert, M. Case, R. A. Dean, A. Hake, K. Sundell, V. Poole Hoffmann, C. Carlson, R. Khanna, M. Mintun, R. DeMattos, K. J. Selzler, E. Siemers, Trial of Solanezumab for Mild Dementia Due to Alzheimer's Disease, N. Engl. J. Med. 378 (2018) 321–330.

[12] M. F. Egan, J. Kost, T. Voss, Y. Mukai, P. S. Aisen, J. L. Cummings, P. N. Tariot, B. Vellas, C. H. van Dyck, M. Boada, Y. Zhang, W. Li, C. Furtek, E. Mahoney, L. Harper Mozley, Y. Mo, C. Sur, D. Michelson, Randomized Trial of Verubecestat for Prodromal Alzheimer's Disease, N. Engl. J. Med. 380 (2019) 1408–1420.

[13] A. M. Wessels, P. N. Tariot, J. A. Zimmer, K. J. Selzler, S. M. Bragg, S. W. Andersen, J. Landry, J. H. Krull, A. M. Downing, B. A. Willis, S. Shcherbinin, J. Mullen, P. Barker, J. Schumi, C. Shering, B. R. Matthews, R. A. Stern, B. Vellas, S. Cohen, E. MacSweeney, M. Boada, J. R. Sims, Efficacy and Safety of Lanabecestat for Treatment of Early and Mild Alzheimer Disease: The AMARANTH and DAYBREAK-ALZ Randomized Clinical Trials, JAMA Neurol (2019).

[14] D. Henley, N. Raghavan, R. Sperling, P. Aisen, R. Raman, G. Romano, Preliminary Results of a Trial of Atabecestat in Preclinical Alzheimer's Disease, N. Engl. J. Med. 380 (2019) 1483–1485.

[15] P. S. Aisen, E. Siemers, D. Michelson, S. Salloway, C. Sampaio, M. C. Carrillo, R. Sperling, R. Doody, P. Scheltens, R. Bateman, M. Weiner, B. Vellas, What Have We Learned from Expedition III and EPOCH Trials? Perspective of the CTAD Task Force, J Prev Alzheimers Dis 5 (2018) 171–174.

[16] G. Klein, P. Delmar, N. Voyle, S. Rehal, C. Hofmann, D. Abi-Saab, M. Andjelkovic, S. Ristic, G. Wang, R. Bateman, G. A. Kerchner, M. Baudler, P. Fontoura, R. Doody, Gantenerumab reduces amyloid-β plaques in patients with prodromal to moderate alzheimer's disease: a pet substudy interim analysis, Alzheimer's Research & Therapy 11 (2019) 101. URL: `https://doi.org/10.1186/s13195-019-0559-z`. doi:`10.1186/s13195-019-0559-z`.

[17] H. M. Fonteijn, M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill, S. J. Tabrizi, S. Ourselin, N. C. Fox, D. C. Alexander, An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease, NeuroImage 60 (2012) 1880–1889.

[18] B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak, B. Caffo, J. L. Prince, A computational neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging Initiative cohort, NeuroImage 63 (2012) 1478–1486.

[19] N. P. Oxtoby, S. Garbarino, N. C. Firth, J. D. Warren, J. M. Schott, D. C. Alexander, Data-Driven Sequence of Changes to Anatomical Brain Connectivity in Sporadic Alzheimer's Disease, Front Neurol 8 (2017) 580.

[20] J. Schiratti, S. Allassonnière, O. Colliot, S. Durrleman, Learning spatiotemporal trajectories from manifold-valued longitudinal data, in: NIPS, 2015, pp. 2404–2412.

[21] C. A. Nader, N. Ayache, P. Robert, M. Lorenzi, A. D. N. Initiative, Monotonic gaussian process for spatio-temporal disease progression modeling in brain imaging data, NeuroImage 205 (2020). URL: `https://doi.org/10.1016/j.neuroimage.2019.116266`. doi:`10.1016/j.neuroimage.2019.116266`.

[22] M. Bilgel, B. Jedynak, D. F. Wong, S. M. Resnick, J. L. Prince, Temporal Trajectory and Progression Score Estimation from Voxelwise Longitudinal Imaging Measures: Application to Amyloid Imaging, Inf Process Med Imaging 24 (2015) 424–436.

[23] M. C. Donohue, H. Jacqmin-Gadda, M. L. Goff, R. G. Thomas, R. Raman, A. C. Gamst, L. A. Beckett, C. R. Jack, M. W. Weiner, J.-F. Dartigues, P. S. Aisen, Estimating long-term multivariate progression from short-term data, Alzheimer's & Dementia 10 (2014) S400 – S410. doi:`https://doi.org/10.1016/j.jalz.2013.10.003`.

[24] Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, J. M. Mateos-P?rez, A. C. Evans, A. D. N. Initiative., Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis, Nat Commun 7 (2016) 11934.

[25] M. Lorenzi, M. Filippone, G. B. Frisoni, D. C. Alexander, S. Ourselin, Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in alzheimer's disease, NeuroImage (2017). doi:`https://doi.org/10.1016/j.Neuroimage.2017.08.059`.

[26] R. V. Marinescu, A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, S. J. Crutch, D. C. Alexander, DIVE: A spatiotemporal progression model of brain pathology in neurodegenerative disorders, NeuroImage 192 (2019) 166–177.

[27] A. L. Young, N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, D. C. Alexander, A data-driven model of biomarker changes in sporadic Alzheimer's disease, Brain 137 (2014) 2564–2577.

[28] N. P. Oxtoby, A. L. Young, D. M. Cash, T. L. S. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, D. C. Alexander, Data-driven models of dominantly-inherited Alzheimer's disease progression, Brain 141 (2018) 1529–1544.

[29] D. Li, S. Iddi, W. K. Thompson, M. C. Donohue, Bayesian latent time joint mixed effect models for multicohort longitudinal data, Stat Methods Med Res 28 (2019) 835–845.

[30] W. Hao, A. Friedman, Mathematical model on Alzheimer's disease, BMC Syst Biol 10 (2016) 108.

[31] J. R. Petrella, W. Hao, A. Rao, P. M. Doraiswamy, Computational Causal Modeling of the Dynamic Biomarker Cascade in Alzheimer's Disease, Comput Math Methods Med 2019 (2019) 6216530.

[32] Y. Iturria-Medina, F. M. Carbonell, R. C. Sotero, F. Chouinard-Decorte, A. C. Evans, Multifactorial causal model of brain (dis)organization and therapeutic intervention: Application to alzheimer's disease, NeuroImage 152 (2017) 60 – 77. URL: http://www.sciencedirect.com/science/article/pii/S1053811917301684. doi:https://doi.org/10.1016/j.neuroimage.2017.02.058.

[33] S. Garbarino, M. Lorenzi, Modeling and Inference of Spatio-Temporal Protein Dynamics Across Brain Networks, in: IPMI 2019 - 26th International Conference on Information Processing in Medical Imaging, volume 11492 of *LNCS*, Springer, Hong-Kong, China, 2019, pp. 57–69. URL: https://hal.inria.fr/hal-02165021.

[34] L. Antelmi, N. Ayache, P. Robert, M. Lorenzi, Sparse Multi-Channel Variational Autoencoder for the Joint Analysis of Heterogeneous Data, in: ICML 2019 - 36th International Conference on Machine Learning, Long Beach, United States, 2019.

[35] L. M. Shaw, H. Vanderstichele, M. Knapik-Czajka, C. M. Clark, P. S. Aisen, R. C. Petersen, K. Blennow, H. Soares, A. Simon, P. Lewczuk, R. Dean, E. Siemers, W. Potter, V. M. Lee, J. Q. Trojanowski, Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects, Ann. Neurol. 65 (2009) 403–413.

[36] D. M. Cash, C. Frost, L. O. Iheme, D. ?nay, M. Kandemir, J. Fripp, O. Salvado, P. Bourgeat, M. Reuter, B. Fischl, M. Lorenzi, G. B. Frisoni, X. Pennec, R. K. Pierson, J. L. Gunter, M. L. Senjem, C. R. Jack, N. Guizard, V. S. Fonov, D. L. Collins, M. Modat, M. J. Cardoso, K. K. Leung, H. Wang, S. R. Das, P. A. Yushkevich, I. B. Malone, N. C. Fox, J. M. Schott, S. Ourselin, Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge, Neuroimage 123 (2015) 149–164.

[37] N. Schuff, N. Woerner, L. Boreta, T. Kornfield, L. M. Shaw, J. Q. Trojanowski, P. M. Thompson, C. R. Jack, M. W. Weiner, MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers, Brain 132 (2009) 1067–1077.

[38] R. Marinescu, A. Eshaghi, D. Alexander, P. Golland, Brainpainter: A software for the visualisation of brain structures, biomarkers and associated pathological processes, arXiv preprint arXiv:1905.08627 (2019).

[39] D. Gamberger, N. Lavrač, S. Srivatsa, R. E. Tanzi, P. M. Doraiswamy, Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease, Sci Rep 7 (2017) 6763.

[40] R. Kochhann, J. S. Varela, C. S. M. Lisboa, M. L. F. Chaves, The Mini Mental State Examination: Review of cutoff points adjusted for schooling in a large Southern Brazilian sample, Dement Neuropsychol 4 (2010) 35–41.

[41] R. A. Sperling, C. R. Jack, P. S. Aisen, Testing the right target and right drug at the right stage, Sci Transl Med 3 (2011) 111cm33.

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

[43] H. Zetterberg, S. C. Burnham, Blood-based molecular biomarkers for alzheimer's disease, Molecular brain 12 (2019) 26–26. doi:10.1186/s13041-019-0448-1.

[44] K. Blennow, H. Hampel, M. Weiner, H. Zetterberg, Cerebrospinal fluid and plasma biomarkers in Alzheimer disease, Nat Rev Neurol 6 (2010) 131–144.

[45] S. Westwood, E. Leoni, A. Hye, S. Lynham, M. R. Khondoker, N. J. Ashton, S. J. Kiddle, A. L. Baird, R. Sainz-Fuertes, R. Leung, J. Graf, C. T. Hehir, D. Baker, C. Cereda, C. Bazenet, M. Ward, M. Thambisetty, S. Lovestone, Blood-Based Biomarker Candidates of Cerebral Amyloid Using PiB PET in Non-Demented Elderly, J. Alzheimers Dis. 52 (2016) 561–572.

[46] S. Gauthier, J. Alam, H. Fillit, T. Iwatsubo, H. Liu-Seifert, M. Sabbagh, S. Salloway, C. Sampaio, J. R. Sims, B. Sperling, R. Sperling, K. A. Welsh-Bohmer, J. Touchon, B. Vellas, P. Aisen, Combination Therapy for Alzheimer's Disease: Perspectives of the EU/US CTAD Task Force, J Prev Alzheimers Dis 6 (2019) 164–168.

[47] S. K. Kaufman, K. Del Tredici, T. L. Thomas, H. Braak, M. I. Diamond, Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in alzheimer's disease and part, Acta Neuropathologica 136 (2018) 57–67. URL: https://doi.org/10.1007/s00401-018-1855-6. doi:10.1007/s00401-018-1855-6.

[48] M. J. Pontecorvo, M. D. Devous, I. Kennedy, M. Navitsky, M. Lu, N. Galante, S. Salloway, P. M. Doraiswamy, S. Southekal, A. K. Arora, A. McGeehan, N. C. Lim, H. Xiong, S. P. Truocchio, A. D. Joshi, S. Shcherbinin, B. Teske, A. S. Fleisher, M. A. Mintun, A multicentre longitudinal study of flortaucipir (18F) in normal ageing, mild cognitive impairment and Alzheimer's disease dementia, Brain 142 (2019) 1723–1735.

[49] J. Cummings, K. Blennow, K. Johnson, M. Keeley, R. J. Bateman, J. L. Molinuevo, J. Touchon, P. Aisen, B. Vellas, Anti-Tau Trials for Alzheimer's Disease: A Report from the EU/US/CTAD Task Force, J Prev Alzheimers Dis 6 (2019) 157–163.

[50] M. Safieh, A. D. Korczyn, D. M. Michaelson, ApoE4: an emerging therapeutic target for Alzheimer's disease, BMC Medicine 17 (2019).

[51] J. Kim, J. M. Basak, D. M. Holtzman, The role of apolipoprotein E in Alzheimer's disease, Neuron 63 (2009) 287–303.

[52] R. Sivera, N. Capet, V. Manera, R. Fabre, M. Lorenzi, H. Delingette, X. Pennec, N. Ayache, P. Robert, Voxel-based assessments of treatment effects on longitudinal brain changes in the multidomain alzheimer preventive trial cohort, Neurobiology of Aging 94 (2020) 50 – 59. URL: http://www.sciencedirect.com/science/article/pii/S0197458020301317. doi:https://doi.org/10.1016/j.neurobiolaging.2019.11.020.

[53] M. Lorenzi, X. Pennec, G. B. Frisoni, N. Ayache, Disentangling normal aging from alzheimer's disease in structural magnetic resonance images, Neurobiology of Aging 36 (2015) S42 – S52. URL: http://www.sciencedirect.com/science/article/pii/S0197458014005594. doi:https://doi.org/10.1016/j.neurobiolaging.2014.07.046, novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD).

[54] M. Reuter, N. J. Schmansky, H. D. Rosas, B. Fischl, Within-subject template estimation for unbiased longitudinal image analysis, NeuroImage 61 (2012) 1402–1418.

[55] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, R. J. Killiany, An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest, NeuroImage 31 (2006) 968–980.

[56] Z. Ghahramani, M. Beal, Graphical models and variational methods, in: M. Opper, D. Saad (Eds.), Advanced mean field methods: theory and practice, Neural Information Processing, MIT, 2001.

[57] D. P. Kingma, M. Welling, Auto-encoding variational bayes, CoRR abs/1312.6114 (2013). arXiv:1312.6114.
[58] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014. URL: http://arxiv.org/abs/1412.6980, cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.
[59] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer, Automatic differentiation in pytorch (2017).
[60] T. Q. Chen, Y. Rubanova, J. Bettencourt, D. K. Duvenaud, Neural ordinary differential equations, in: S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett (Eds.), Advances in Neural Information Processing Systems 31, Curran Associates, Inc., 2018, pp. 6571–6583. URL: http://papers.nips.cc/paper/7892-neural-ordinary-differential-equations.pdf.
[61] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, Springer New York, New York, NY, 1998, pp. 199–213. URL: https://doi.org/10.1007/978-1-4612-1694-0_15. doi:10.1007/978-1-4612-1694-0_15.

Acknowledgments

This work has been supported by the French government, through the UCA^{JEDI} and 3IA Côte d'Azur Investments in the Future project managed by the National Research Agency (ref.n ANR-15-IDEX-01 and ANR-19-P3IA-0002), the grant AAP Santé 06 2017-260 DGA-DSH, and by the Inria Sophia-Antipolis-Méditerranée, "NEF" computation cluster.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) and DOD ADNI. 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.

Author information

Clément Abi Nader and Marco Lorenzi designed the method. Implementation was carried out by Clément Abi Nader. The manuscript was written by Clément Abi Nader with support from Marco Lorenzi, Nicholas Ayache, Giovanni Frisoni, and Philippe Robert.

Competing interests

The authors declare no competing interests.

Supplementary Material

1) Time-shift comparison and validation

We compared our estimated disease severity (Figure 3 in the manuscript) with the one obtained applying the monotonic Gaussian Process (GP) model of [25] from the state-of-the-art (Figure 1A). While both methods estimate significant time differences when going from healthy to pathological stages, our approach captures a larger temporal variability for both earlier and later stages of the disease, as shown in Figure 1B, highlighting a stronger separability across clinical stages.

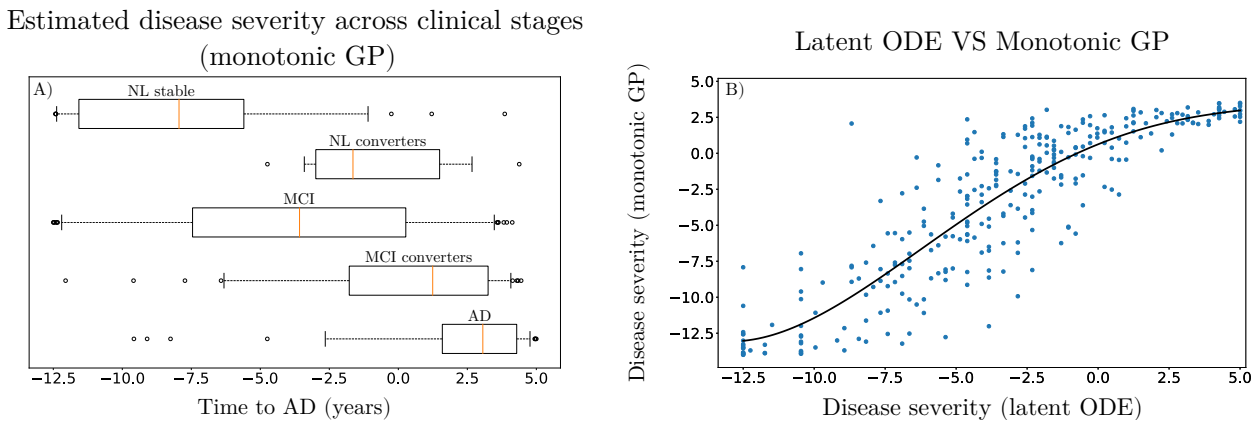

Figure 1: A: Distribution of the disease severity estimated by the monotonic GP method [25] on the training set. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer's dementia. B: Comparison of the disease severity estimated by our method (denoted as latent ODE) with respect to the one estimated by the monotonic GP.

We also assessed the model on an independent testing cohort from the ADNI composed of 130 NL stable, 10 NL converters, 125 MCI stable, 7 MCI converters, and 12 AD subjects which were not necessarily amyloid positive. It is important to note that no PET-FDG data was available for these subjects. We provide in Table 1 socio-demographic and clinical information for the testing cohort across the different clinical groups. Despite the fact that no FDG data was used to estimate the disease severity, we observe in Figure 2 that the method still exhibits good separating performances between clinical stages, coherently with the clinical status of the testing individuals.

Table 1: Baseline socio-demographic and clinical information for testing cohort (284 subjects for 2116 data points). Average values, standard deviation in parenthesis. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer's dementia. ADAS11: Alzheimer's Disease Assessment Scale-cognitive subscale, 11 items. AV45: (18)F-florbetapir Amyloid PET imaging. SUVR: Standardized Uptake Value Ratio.

	NL stable	NL converters	MCI stable	MCI converters	AD
N	130	10	125	7	12
Age (yrs)	72 (6)	74 (8)	71 (8)	73 (9)	78 (6)
Education (yrs)	17 (2)	16 (2)	16 (3)	14 (3)	17 (2)
ADAS11	5.4 (2.8)	7.7 (4.1)	7.8 (3.3)	14.3 (5.2)	15.0 (6.7)
WholeBrain (cm^3)	1063 (103)	1104 (98)	1054 (97)	966 (104)	1010 (108)
AV45 (SUVR)	0.9 (0.1)	1.0 (0.1)	1.0 (0.1)	1.1 (0.2)	1.2 (0.3)

2) Simulated clinical endpoints

We provide in Table 2 the estimated values for each clinical score at predicted conversion time for the normal progression case when performing the simulations presented in Section 2.4.

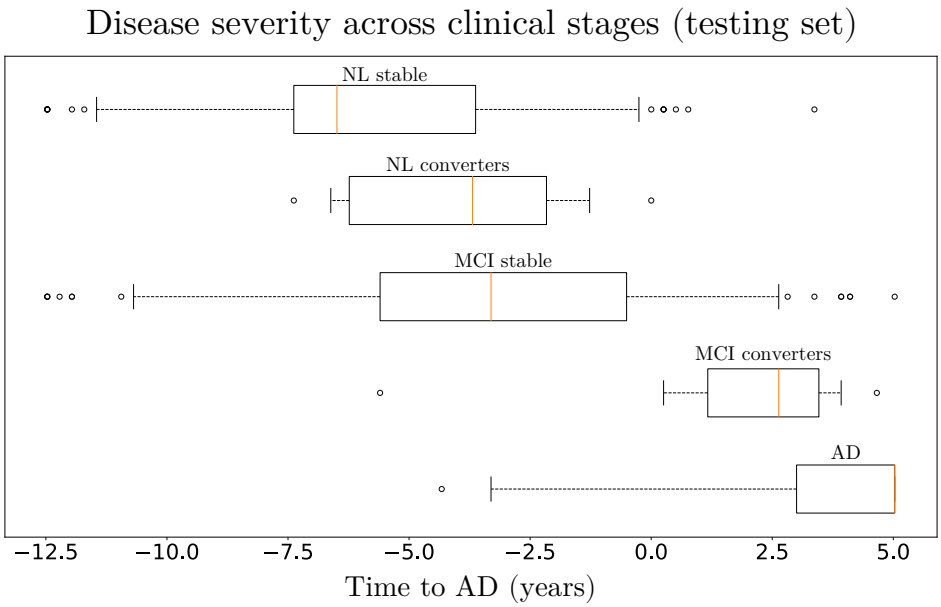

Figure 2: Distribution of the disease severity estimated for the subjects of the testing set, relatively to the long-term dynamics of Figure 3 in the manuscript. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer's dementia.

Table 2: Estimated mean (standard deviation) of the clinical outcomes at predicted conversion time for the normal progression case by year of simulated intervention (100% and 50% amyloid lowering interventions). Results in bold indicate a statistically significant difference between placebo and treated scenarios (p<0.01, two-sided t-test, 100 cases per arm). AD: Alzheimer's dementia, ADAS11: Alzheimer's Disease Assessment Scale, MMSE: Mini-Mental State Examination, FAQ: Functional Assessment Questionnaire, RAVLT: Rey Auditory Verbal Learning Test.

Amyloid lowering intervention 100%

Score per intervention time

Score \ Years to AD	−12.5	−10	−7.5	−5	−4	−3	−2	−1
ADAS11	**10.7 (7.5)**	**13.4 (6.2)**	**15.7 (5.3)**	17.0 (4.8)	17.3 (4.7)	17.5 (4.6)	17.6 (4.5)	17.7 (4.5)
MMSE	**26.8 (3.0)**	**25.8 (2.5)**	**24.9 (2.2)**	24.3 (2.0)	24.2 (1.9)	24.1 (1.9)	24.0 (1.8)	24.0 (1.9)
FAQ	**4.6 (5.9)**	**6.7 (4.9)**	**8.4 (4.2)**	9.5 (3.8)	9.8 (3.7)	10.0 (3.6)	10.1 (3.6)	10.1 (3.5)
RAVLT immediate	**35.1 (11.0)**	**31.3 (9.0)**	**28.1 (7.7)**	26.2 (7.0)	25.7 (6.7)	25.4 (6.6)	25.2 (6.5)	25.1 (6.4)
RAVLT forgetting (%)	**63.2 (26.7)**	**71.8 (22.0)**	**79.1 (18.5)**	83.8 (16.4)	85.0 (15.8)	85.8 (15.3)	86.4 (15.0)	86.6 (14.7)

Amyloid lowering intervention 50%

Score per intervention time

Score \ Years to AD	−12.5	−10	−7.5	−5	−4	−3	−2	−1
ADAS11	**14.1 (5.7)**	**15.5 (5.1)**	16.6 (4.8)	17.2 (4.6)	17.4 (4.5)	17.5 (4.5)	17.6 (4.5)	17.6 (4.5)
MMSE	**25.5 (2.2)**	**25.0 (2.0)**	24.5 (1.9)	24.2 (1.8)	24.1(1.8)	24.1 (1.8)	24.1 (1.8)	24.1 (1.8)
FAQ	**7.2 (4.4)**	**8.3 (4.0)**	9.2 (3.8)	9.7 (3.6)	9.8 (3.5)	9.9 (3.5)	10.0 (3.5)	10.0 (3.5)
RAVLT immediate	**30.3 (8.2)**	**28.4 (7.4)**	26.8 (6.9)	25.9 (6.6)	25.6 (6.5)	25.5 (6.4)	25.4 (6.3)	25.3 (6.3)
RAVLT forgetting (%)	**74.8 (19.7)**	**79.2 (17.5)**	82.8 (16.1)	85.1 (16.2)	85.7 (14.9)	86.0 (14.7)	86.3 (14.5)	86.5 (17.4)

3) Lower bound

We provide here the detailed derivation to obtain the ELBO of Equation 6 in the main manuscript.

$$
\begin{aligned}
\log p(\mathbf{X}|\boldsymbol{\sigma}^2, \boldsymbol{\psi}) &= \log\Big[\int p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) p(\mathbf{z}) d\mathbf{z}\Big] \\
&= \log\Big[\int p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) p(\mathbf{z}) \frac{q(\mathbf{z}|\mathbf{X})}{q(\mathbf{z}|\mathbf{X})} d\mathbf{z}\Big] \\
&= \log\Big[\mathbb{E}_{q(\mathbf{z}|\mathbf{X})} \frac{p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) p(\mathbf{z})}{q(\mathbf{z}|\mathbf{X})}\Big] \\
&\overset{\text{Jensen}}{\geq} \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\Big[\log \frac{p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) p(\mathbf{z})}{q(\mathbf{z}|\mathbf{X})}\Big] \\
&= \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\Big[\log p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi})\Big] - \mathcal{D}\Big[q(\mathbf{z}|\mathbf{X})|p(\mathbf{z})\Big] \\
&= \mathcal{E}.
\end{aligned} \tag{1}
$$

Given that:

$$
p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma}^2, \boldsymbol{\psi}) = \prod_m p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m),\ q(\mathbf{z}|\mathbf{X}) = \prod_m q(z^m|\mathbf{X}),\ \text{and},\ p(\mathbf{z}) = \mathcal{N}(\mathbf{0}, \mathbf{I}).
$$

We obtain:

$$
\mathcal{E} = \sum_m E_{q(\mathbf{z}|\mathbf{X})}\Big[\log p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m)\Big] - \mathcal{D}\Big[q(z^m|\mathbf{x}^m)|p(z^m)\Big]. \tag{2}
$$

4) KL divergence

We have that:

$$
\begin{aligned}
q(z^m|\mathbf{X}) &= \mathcal{N}(f(\mathbf{x}^m, \phi_m^1), h(\mathbf{x}^m, \phi_m^2)), \\
p(z^m) &= \mathcal{N}(0, 1).
\end{aligned} \tag{3}
$$

We use the closed-form formula to calculate the KL divergence between two normal distributions:

$$
\begin{aligned}
\mathcal{D}\Big[q(\mathbf{z}|\mathbf{X})|p(\mathbf{z})\Big] &= \sum_m \mathcal{D}\Big[q(z^m|\mathbf{x}^m)|p(z^m)\Big] \\
&= \frac{1}{2}\sum_m \Big[-\log(h(\mathbf{x}^m, \phi_m^2)) - 1 + h(\mathbf{x}^m, \phi_m^2) + f(\mathbf{x}^m, \phi_m^1)^2\Big]
\end{aligned} \tag{4}
$$

5) Graphical model

Figure 3 below provides the graphical model illustrating the method presented in Section 4.

6) ELBO computation

Algorithm 1 below details the steps to compute the ELBO for a given subject i at time t.

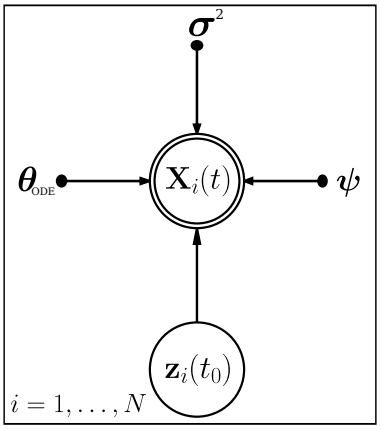

Figure 3: Graphical model of the proposed method.

Algorithm 1 Forward pass to compute the ELBO for a given subject i at time t.

function COMPUTE_ELBO($\mathbf{X}(t), \mathbf{X}(t_0), \theta_{ODE}, \boldsymbol{\psi}, \boldsymbol{\phi}, \boldsymbol{\sigma}^2$)
For ease of notation we drop the i index in the pseudo-code.
Sample $\mathbf{z}(t_0) \sim q(\mathbf{z}(t_0)|\mathbf{X}(t_0)) = \prod_m \mathcal{N}(f(\mathbf{x}^m(t_0), \phi_m^1), h(\mathbf{x}^m(t_0), \phi_m^2))$ ▹ Baseline latent representation (reparameterization trick).
Compute $\mathbf{z}(t) = MIDPOINT(\mathbf{z}(t_0), g, \theta_{ODE}, t)$ ▹ Predict latent representation at time t by numerically solving the ODEs system.
Compute $\mathbb{E}_{q(\mathbf{z}(t_0)|\mathbf{X}(t_0))}\Big[\log p(\mathbf{x}^m(t)|\mathbf{z}(t), \theta_{ODE}, \sigma_m^2, \psi_m)\Big] \approx -\frac{D_m}{2}\log(2\pi\sigma_m^2) - \frac{1}{2\sigma_m^2}||\mathbf{x}^m(t) - \mu_m(\mathbf{z}(t))||^2$ ▹ Expectation term Equation 6.
Compute $\mathcal{D}\Big[q(z^m(t_0)|\mathbf{x}^m(t_0))|p(z^m(t_0))\Big] = \frac{1}{2}\Big[-\log(h(\mathbf{x}^m(t_0), \phi_m^2)) - 1 + h(\mathbf{x}^m(t_0), \phi_m^2) + f(\mathbf{x}^m(t_0), \phi_m^1)^2\Big]$ ▹ KL divergence Equation 6.
Compute $\mathcal{E} = \sum_m \mathbb{E}_{q(\mathbf{z}(t_0)|\mathbf{X}(t_0))}\Big[\log p(\mathbf{x}^m(t)|\mathbf{z}(t), \theta_{ODE}, \sigma_m^2, \psi_m)\Big] - \mathcal{D}\Big[q(z^m(t_0)|\mathbf{x}^m(t_0))|p(z^m(t_0))\Big]$.
Return $\mathcal{E}$
end function