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

Clément Abi Nader[1*], **Nicholas Ayache**[1], **Giovanni B. Frisoni**[2], **Philippe Robert**[3], **Marco Lorenzi**[1], **for the Alzheimer's Disease Neuroimaging Initiative**

*For correspondence:
clement.abi-nader@inria.fr

[1]Université Côte d'Azur, Inria Sophia Antipolis, Epione Research Project, France; [2]Memory Clinic and LANVIE-Laboratory of Neuroimaging of Aging, Hospitals and University of Geneva, Geneva, Switzerland; [3]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.

Introduction

The number of people affected by Alzheimer's Disease (AD) has recently exceeded 46 millions and is expected to double every 20 years (*Prince et al., 2015*), 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 (*Jack and Holtzman, 2013*; *Jack et al., 2013*). 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" (*Villemagne et al., 2013*; *Murphy and LeVine, 2010*; *Delacourte et al., 1999*; *Braak and Braak, 1991*; *Bateman et al., 2012*). Based on this rationale, clinical trials have been focusing on the development and testing of disease modifiers

40 targeting amyloid β aggregates (*Cummings et al., 2019b*), for example by increasing its clearance
41 or blocking its accumulation. Although the amyloid hypothesis has been recently invigorated by a
42 post-hoc analysis of the aducanumab trial (*Howard and Liu, 2020*), clinical trials failed so far to show
43 efficacy of this kind of treatments, as the clinical primary endpoints were not met (*Honig et al., 2018*;
44 *Egan et al., 2019*; *Wessels et al., 2019*), or because of unacceptable adverse effects (*Henley et al.,*
45 *2019*). In the past years, growing consensus emerged about the critical importance of intervention
46 time, and about the need of starting anti-amyloid treatments during the pre-symptomatic stages
47 of the disease (*Aisen et al., 2018*). Nevertheless, the design of optimal intervention strategies is
48 currently not supported by quantitative analysis methods allowing to model and assess the effect
49 of intervention time and dosing (*Klein et al., 2019*). The availability of models of the pathophysi-
50 ology of AD would entail great potential to test and analyze clinical hypothesis characterizing AD
51 mechanisms, progression, and intervention scenarios.
52
53 Within this context, quantitative models of disease progression, referred to as Disease Progression
54 Models (DPMs), have been proposed (*Fonteijn et al., 2012*; *Jedynak et al., 2012*; *Oxtoby et al., 2017*;
55 *Schiratti et al., 2015*; *Abi Nader et al., 2020*), to quantify the dynamics of the changes affecting
56 the brain during the whole disease span. These models rely on the statistical analysis of large
57 datasets of different data modalities, such as clinical scores, or brain imaging measures derived
58 from Magnetic Resonance Imaging (MRI), Amyloid- and Fluorodeoxyglucose-Positron Emission
59 Tomography (PET) (*Bilgel et al., 2015*; *Donohue et al., 2014*; *Iturria-Medina et al., 2016*). In general,
60 DPMs estimate a long-term disease evolution from the joint analysis of multivariate time-series
61 acquired on a short-term time-scale. Due to the temporal delay between the disease onset and
62 the appearance of the first symptoms, DPMs rely on the identification of an appropriate tempo-
63 ral reference to describe the long-term disease evolution (*Lorenzi et al., 2017*; *Marinescu et al.,*
64 *2019a*). These tools are promising approaches for the analysis of clinical trials data, as they allow
65 to represent the longitudinal evolution of multiple biomarkers through a global model of disease
66 progression. Such a model can be subsequently used as a reference in order to stage subjects and
67 quantify their relative progression speed (*Young et al., 2014*; *Oxtoby et al., 2018*; *Li et al., 2019*).
68 However, these approaches remain purely descriptive as they don't account for causal relationships
69 among biomarkers. Therefore, they generally don't allow to simulate progression scenarios based
70 on hypothetical intervention strategies, thus providing a limited interpretation of the pathological
71 dynamics. This latter capability is of utmost importance for planning and assessment of disease
72 modifying treatments.
73
74 To fill this gap, recent works such as (*Hao and Friedman, 2016*; *Petrella et al., 2019*) proposed to
75 model AD progression based on specific assumptions on the biochemical processes of pathologi-
76 cal protein propagation. These approaches explicitly define biomarkers interactions through the
77 specification of sets of Ordinary Differential Equations (ODEs), and are ideally suited to simulate
78 the effect of drug interventions (*Iturria-Medina et al., 2017*). However, these methods are mostly
79 based on the arbitrary choices of pre-defined evolution models, which are not inferred from data.
80 This issue was recently addressed by (*Garbarino and Lorenzi, 2019*), where the authors proposed
81 an hybrid modeling method combining traditional DPMs with dynamical models of AD progres-
82 sion. Still, since this approach requires to design suitable models of protein propagation across
83 brain regions, extending this method to jointly account for spatio-temporal interactions between
84 several processes, such as amyloid propagation, glucose hypometabolism, and brain atrophy, is
85 considerably more complex. Finally, these methods are usually designed to account for imaging
86 data only, which prevents to jointly simulate heterogeneous measures (*Antelmi et al., 2019*), such
87 as image-based biomarkers and clinical outcomes, the latter remaining the reference markers for
88 patients and clinicians.
89
90 In this work we present a novel computational model of AD progression allowing to simulate

Manuscript submitted to eLife

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

Results

109 In the following sections, healthy individuals will be denoted as NL stable, subjects with mild
110 cognitive impairment as MCI stable, subjects diagnosed with Alzheimer's dementia as AD, sub-
111 jects progressing from NL to MCI as NL converters, and subjects progressing from MCI to AD as
112 MCI converters. Amyloid concentration and glucose metabolism are respectively measured by
113 (18)F-florbetapir Amyloid (AV45)-PET and (18)F-fluorodeoxyglucose (FDG)-PET imaging. Cognitive
114 and functional abilities are assessed by the following neuro-psychological tests: Alzheimer's Dis-
115 ease Assessment Scale (ADAS11), Mini-Mental State Examination (MMSE), Functional Assessment
116 Questionnaire (FAQ), Rey Auditory Verbal Learning Test (RAVLT) immediate, and RAVLT forgetting.

Study cohort and biomarkers' changes across clinical groups

118 Our study is based on a cohort of 311 amyloid positive individuals composed of 46 NL stable
119 subjects, 10 NL converters subjects, 106 subjects diagnosed with MCI, 76 MCI converters subjects,
120 and 73 AD patients. The term "amyloid positive" refers to subjects whose amyloid level in the
121 cerebrospinal fluid (CSF) was below the nominal cutoff of 192 pg/ml (*Shaw et al., 2009*) either at
122 baseline, or during any follow-up visit, and conversion to AD was determined using the last available
123 follow-up information. The length of follow-up varies between subjects and goes from 0 to 6 years.
124 Further information about the data are available on https://adni.bitbucket.io/reference/, while
125 details on data acquisition and processing are provided in Section Data acquisition and prepro-
126 cessing. We show in *Table 1A* socio-demographic information for the training cohort across the
127 different clinical groups. *Table 1B* shows baseline values and annual rates of change across clinical
128 groups for amyloid burden (average normalized AV45 uptake in frontal cortex, anterior cingulate,
129 precuneus and parietal cortex), glucose hypometabolism (average normalized FDG uptake in frontal
130 cortex, anterior cingulate, precuneus and parietal cortex), for hippocampal and medial temporal
131 lobe volumes, and for the cognitive ability as measured by ADAS11. Compatibly with previously
132 reported results (*Cash et al., 2015*; *Schuff et al., 2009*), we observe that while regional atrophy,
133 hypometabolism and cognition show increasing rate of change when moving from healthy to
134 pathological conditions, the change of AV45 is maximum in NL stable and MCI stable subjects. We
135 also notice the increased magnitude of ADAS11 in AD as compared to the other clinical groups.
136 Finally, the magnitude of change of FDG is generally milder than the atrophy rates.

137

138 The observations presented in *Table 1* provide us with a glimpse into the biomarkers' trajectories
139 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).

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]

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

167 z-score z^{cli} is derived from neuro-psychological scores: ADAS11, MMSE, FAQ, RAVLT immediate,
168 and RAVLT forgetting. Further details about experimental setup, method formulation, and data
169 pre-processing are given in Section Methods.
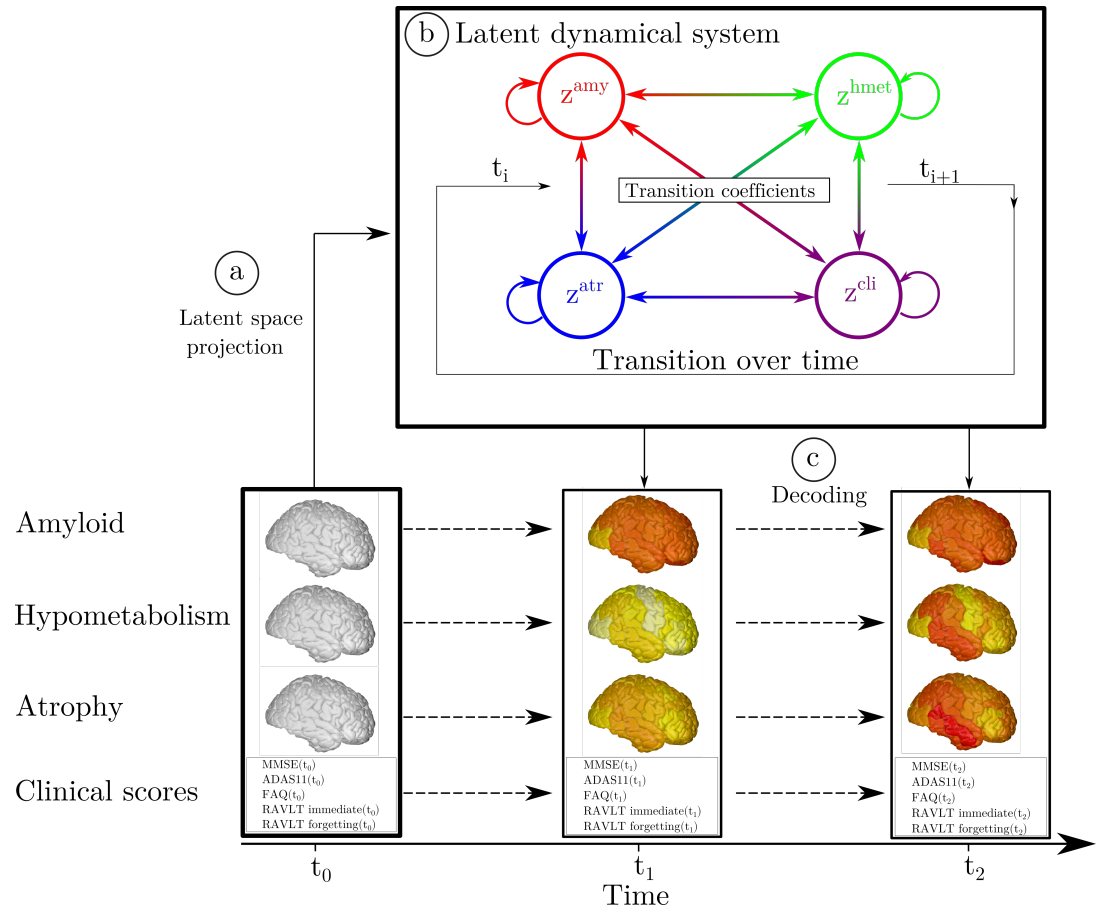
Figure 1. Overview of the method. a) High-dimensional multi-modal measures are projected into a 4-dimensional latent space. Each data modality is transformed in a corresponding z-score $z^{amy}, z^{hmet}, z^{atr}, z^{cli}$. b) The dynamical system describing the relationships between the z-scores allows 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.

Progression model and latent relationships

171 We show in *Figure 2* the dynamical relationships across the different z-scores estimated by the
172 model, where direction and intensity of the arrows quantify the estimated increase of one variable
173 with respect to the other. Being the scores adimensional, they have been conveniently rescaled to
174 the range [0,1] indicating increasing pathological levels. These relationships extend the summary
175 statistics reported in *Table 1* to a much finer temporal scale and wider range of possible biomarkers'
176 values. We observe in *Figure 2A*, *Figure 2B* and *Figure 2C* that large values of the amyloid score z^{amy}
177 trigger the increase of the remaining ones: z^{hmet}, z^{atr}, and z^{cli}. *Figure 2D* shows that large increase
178 of the atrophy score z^{atr} is associated to higher hypometabolism indicated by large values of z^{hmet}.
179 Moreover, we note that high z^{hmet} values also contribute to an increase of z^{cli} (*Figure 2E*). Finally,
180 *Figure 2F* shows that high atrophy values lead to an increase mostly along the clinical dimension z^{cli}.
181 This chain of relationships is in agreement with the cascade hypothesis of AD (*Jack and Holtzman,*
182 *2013*; *Jack et al., 2013*).
183
184 Relying on the dynamical relationships shown in *Figure 2*, starting from any initial set of biomarkers

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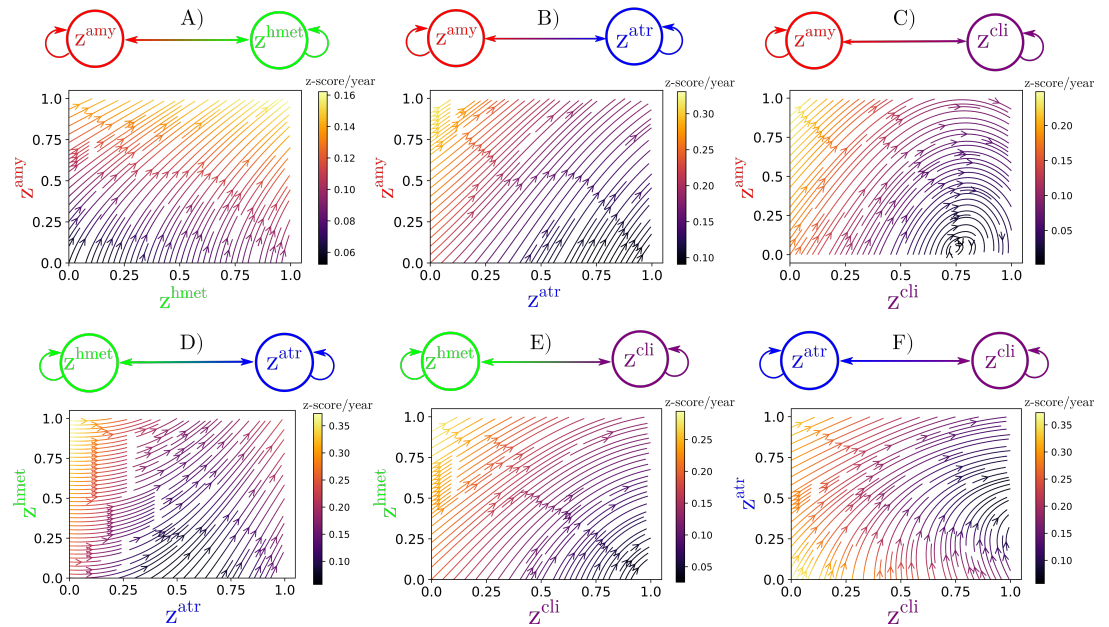
Figure 2. Estimated dynamical relationships across the different z-scores (A to F). Given the values of two z-scores, the arrow at the corresponding coordinates indicates how one score evolves with respect to the other. The intensity of the arrow gives the strength of the relationship between the two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

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 Evaluating disease severity). 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 (*Lorenzi et al., 2017*). The results are provided in *Appendix 1* and show positive generalization results as well as a favourable comparison with the benchmark method.

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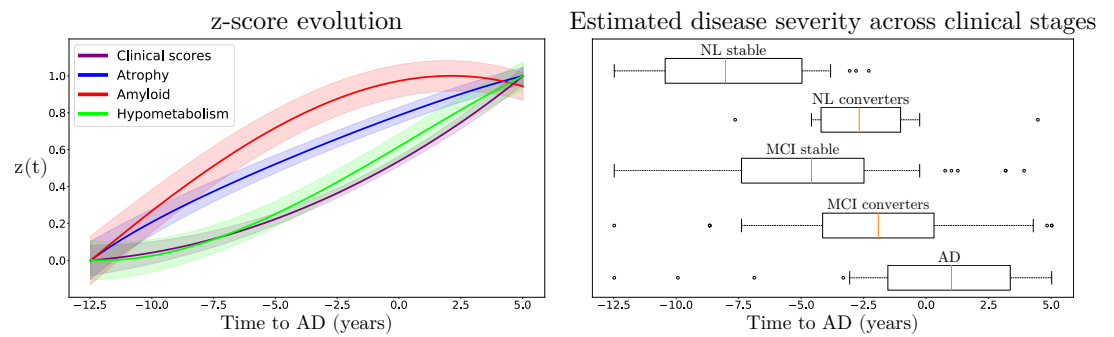
Figure 3. Left: Estimated long-term latent dynamics (time is relative to conversion to Alzheimer's dementia). Shadowed areas represent the standard deviation of the average trajectory. Right: Distribution of the estimated disease severity across clinical stages, relatively to the long-term dynamics on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

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.

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 *Appendix 2 Table 1*) (*Gamberger et al., 2017*; *Kochhann et al., 2010*).
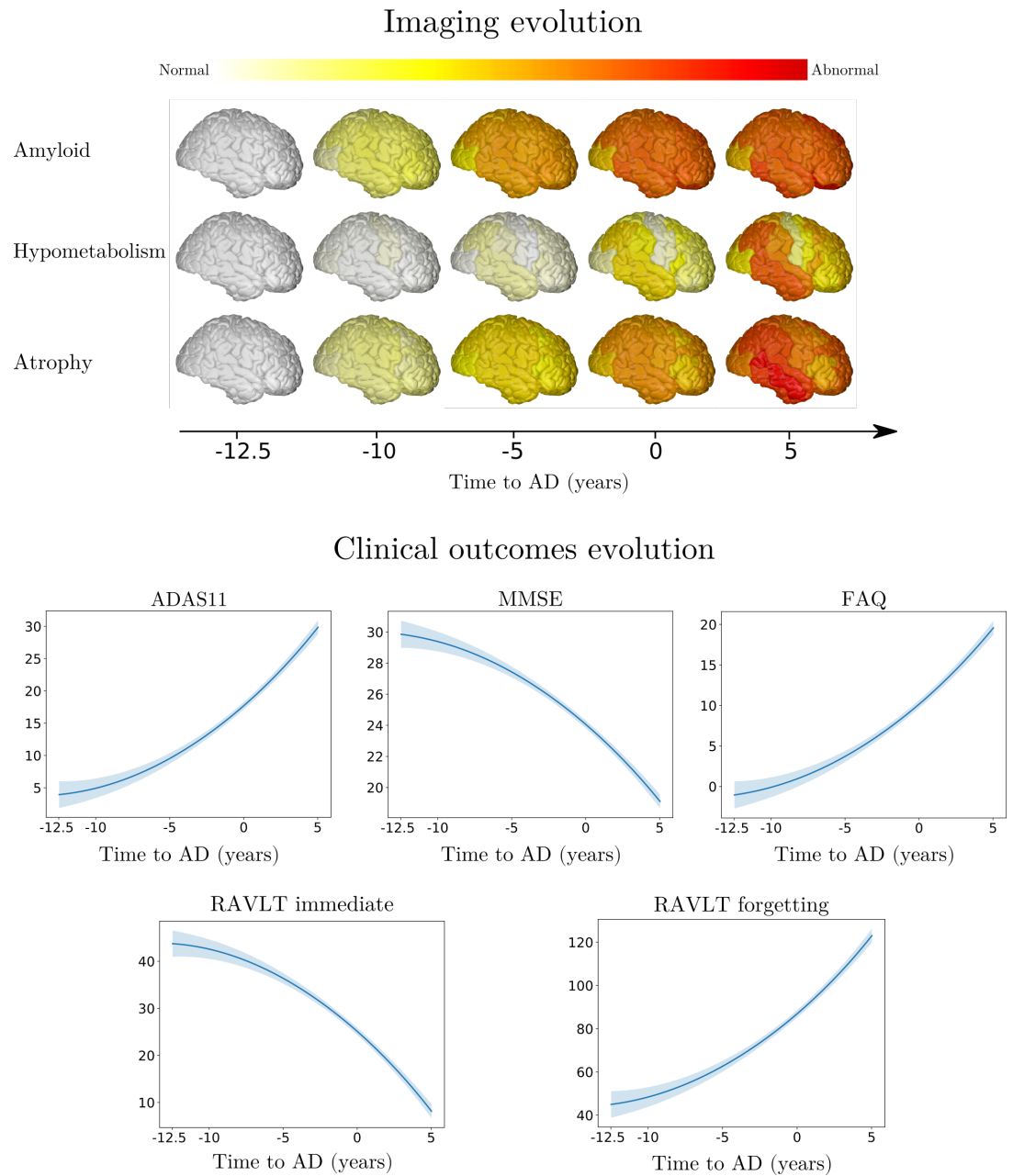
Figure 4. Modelled long-term evolution of cortical measurements for the different types of imaging markers, and clinical scores. Shadowed areas represent the standard deviation of the average trajectory. Brain images were generated using the software provided in (*Marinescu et al., 2019b*).
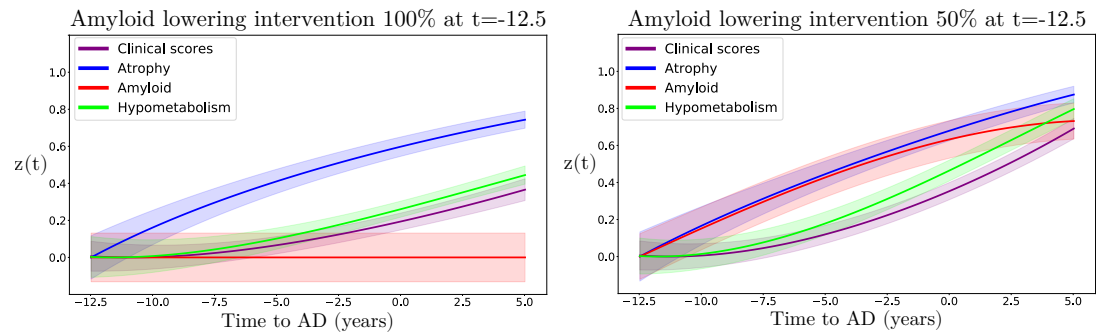
Figure 5. Hypothetical scenarios of irreversible amyloid lowering interventions at *t*=-12.5 years from Alzheimer's dementia diagnosis, with a rate of 100 % (left) or 50 % (right). Shadowed areas represent the standard deviation of the average trajectory.

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

Years to AD / Score	−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

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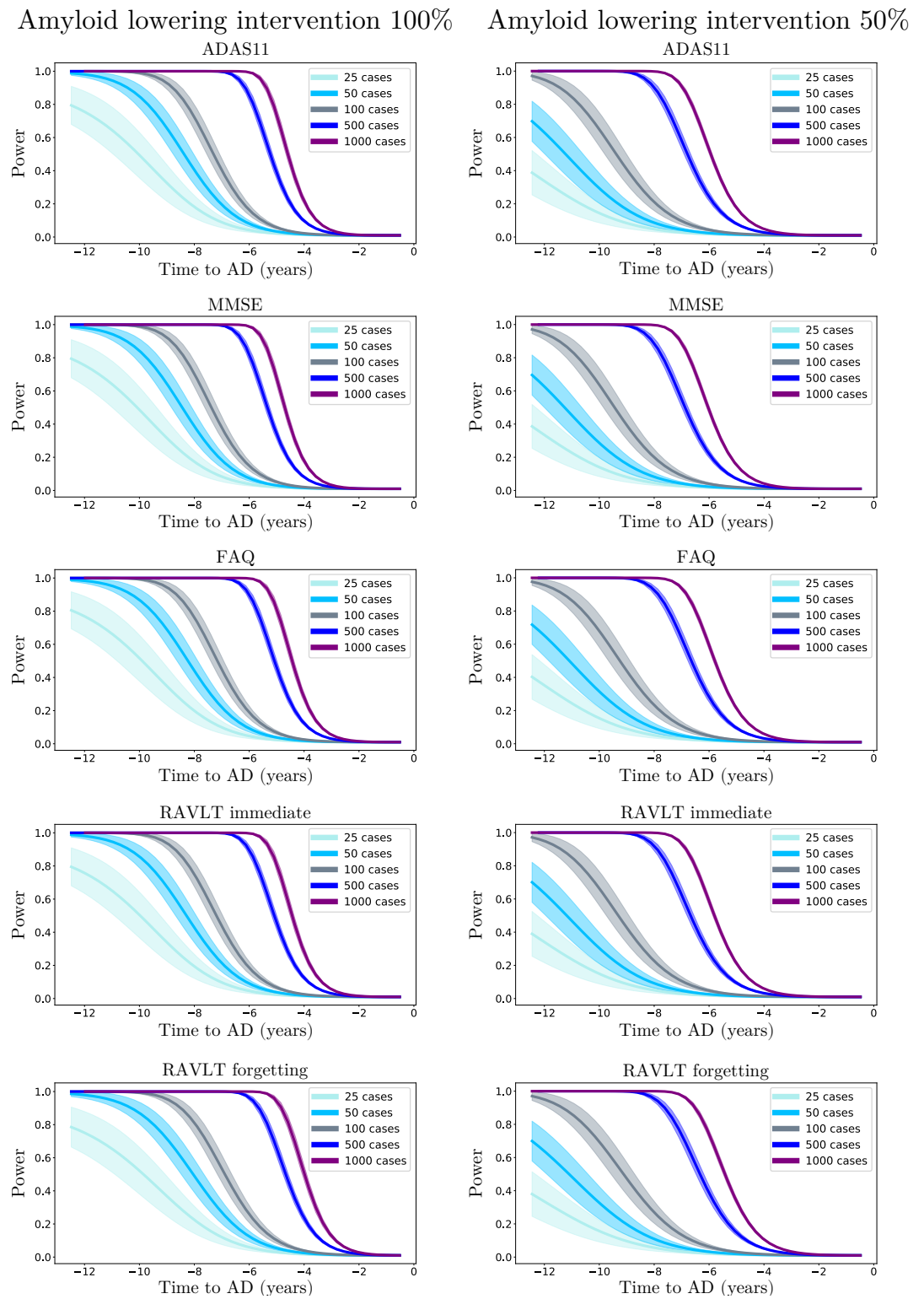
Figure 6. Statistical power of the Student t-test comparing the estimated clinical outcomes at conversion time between placebo and treated scenarios, according to the year of simulated intervention (100% and 50% amyloid lowering) and sample size.

Discussion

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

250 The results obtained with our model are compatible with findings reported in recent clinical studies
251 (*Honig et al., 2018*; *Egan et al., 2019*; *Wessels et al., 2019*). For example, if we consider 500 patients
252 per arm and perform a 100% amyloid lowering intervention for 2 years to reproduce the conditions
253 of the recent trial of Verubecestat (*Egan et al., 2019*), the average improvement of MMSE predicted
254 by our model is of 0.02, falling in the 95% confidence interval measured during that study ([-0.5 ;
255 0.8]). While recent anti-amyloid trials such as (*Honig et al., 2018*; *Egan et al., 2019*; *Wessels et al.,*
256 *2019*) included between 500 and 1000 mild AD subjects per arm and were conducted over a period
257 of two years at most, our analysis suggests that clinical trials performed with less than 1000 subjects
258 with mild AD may be consistently under-powered. Indeed, we see in *Figure 6* that with a sample
259 size of 1000 subjects per arm and a total blockage of amyloid production, a power of 0.8 can be
260 obtained only if intervention is performed at least 5 years before conversion.

263 These results allow to quantify the crucial role of intervention time, and provide an experimental
264 justification for testing amyloid modifying drugs in the pre-clinical stage (*Aisen et al., 2018*; *Sperling*
265 *et al., 2011*). This is for example illustrated in *Table 2*, in which we notice that clinical endpoints are
266 close to placebo even when the simulated intervention takes place up to 5 years before conversion,
267 while stronger cognitive and functional changes happen when amyloid is lowered by 100% or 50%
268 at least 10 years before conversion. These findings may be explained by considering that amyloid
269 accumulates over more than a decade, and that when amyloid clearance occurs the pathological
270 cascade is already entrenched (*Rowe et al., 2010*). Our results are thus supporting the need to iden-
271 tify subjects at the pre-clinical stage, that is to say still cognitively normal, which is a challenging task.
272 Currently, one of the main criteria to enroll subjects into clinical trials is the presence of amyloid in
273 the brain, and blood-based markers are considered as potential candidates for identifying patients
274 at risk for AD (*Zetterberg and Burnham, 2019*). Moreover, recent works such as (*Blennow et al.,*
275 *2010*; *Westwood et al., 2016*) have proposed more complex entry criteria to constitute cohorts
276 based on multi-modal measurements. Within this context, our model could also be used as an
277 enrichment tool by quantifying the disease severity based on multi-modal data as shown in *Figure 3*.
278 Similarly, the method could be applied to predict the evolution of single patient given its current
279 available measurements.

281 An additional critical aspect of anti-amyloid trials is the effect of dose exposure on the production
282 of amyloid (*Klein et al., 2019*). Currently, β-site amyloid precursor protein cleaving enzyme (BACE)
283 inhibitors allow to suppress amyloid production from 50% to 90%. In this study we showed that
284 lowering amyloid by 50% consistently decreases the treatment effect compared to a 100% lowering
285 at the same time. For instance, if we consider a sample size of 1000 subjects per arm in the case of a
286 50% amyloid lowering intervention, an 80% power can be reached only 6.5 years before conversion
287 instead of 5 years for a 100% amyloid lowering intervention. This ability of our model to control the
288 rate of amyloid progression is fundamental in order to provide realistic simulations of anti-amyloid
289 trials.

291 In *Figure 2* we showed that amyloid triggers the pathological cascade affecting the other mark-

ers, 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 (*Gauthier et al., 2019*). For instance, pathologists have shown tau deposition in brainstem nuclei in adolescents and children (*Kaufman et al., 2018*), and clinicians are currently investigating the pathological effect of early tau spreading on AD progression (*Pontecorvo et al., 2019*), raising crucial questions about its relationship with amyloid accumulation, and the impact on cognitive impairment (*Cummings et al., 2019a*). 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 (*Safieh et al., 2019*). 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 (*Kim et al., 2009*). 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 Study cohort and biomarkers' changes across clinical groups). 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 (*Sivera et al., 2020*; *Lorenzi et al., 2015*). 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.

Methods

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 (*Reuter et al., 2012*), 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 (*Desikan et al., 2006*). 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 Study cohort and biomarkers' changes across clinical groups.

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:

$$\frac{d\mathbf{z}(t)}{dt} = \mathbf{W}\mathbf{z}(t) - \mathbf{V}\mathbf{z}^2(t) = g(\mathbf{z}(t), \theta_{ODE}) \text{ where,}$$

$$\left(\mathbf{W}_{i,j}\right) = \begin{cases} k_i & \text{if i=j,} \\ \alpha_{i,j} & \text{otherwise;} \end{cases} \text{ and } \left(\mathbf{V}_{i,j}\right) = \begin{cases} k_i & \text{if i=j} \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

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, $\mathbf{z}(t) = \mathbf{z}(0) + \int_0^t g(\mathbf{z}(x), \theta_{ODE}) dx$.

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:

$$\mathcal{L} = \sum_i^N \log\left[p(\mathbf{X}_i(t)|\theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi})\right]$$

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

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 $q(\mathbf{z}|\mathbf{X})$ (*Ghahramani and Beal, 2001*), in order to derive a lower-bound $\mathcal{E}$ for the marginal log-likelihood:

$$\log p(\mathbf{X}|\theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\left[\log p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \boldsymbol{\sigma^2}, \boldsymbol{\psi})\right]$$

$$- \mathcal{D}\left[q(\mathbf{z}|\mathbf{X})|p(\mathbf{z})\right] \tag{5}$$

$$= \mathcal{E},$$

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:

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

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

Model optimization

Using the reparameterization trick (*Kingma and Welling, 2013*), 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 *Equation 3*. The model is trained by computing the total ELBO for all the subjects at all the available time points. The parameters $\psi, \phi^1, \phi^2, \theta_{ODE}, \sigma$ 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}:

$$
\int_{t_i}^{t_{i+1}} g(\mathbf{z}(x))dx \approx h \cdot g\Big(\mathbf{z}(\frac{t_i + t_{i+1}}{2})\Big)
$$
$$
\approx h \cdot g\Big(\mathbf{z}(t_i) + \frac{h}{2}g(\mathbf{z}(t_i))\Big), \quad h = t_{i+1} - t_i.
$$
(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 ϕ^1 and ϕ^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 (*Kingma and Ba, 2014*) 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 (*Paszke et al., 2017*), and we used the *torchdiffeq* package developed in (*Chen et al., 2018*) to backpropagate through the ODE solver.

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 (*Akaike, 1998*). 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.

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 Variational inference 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:

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

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

432 **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:

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

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

460 **Competing interests**

461 The authors declare no competing interests.

462 **References**

463 **Abi Nader C**, Ayache N, Robert P, Lorenzi M, Initiative ADN. Monotonic Gaussian Process for spatio-temporal
464 disease progression modeling in brain imaging data. NeuroImage. 2020; 205. https://doi.org/10.1016/j.
465 neuroimage.2019.116266, doi: 10.1016/j.neuroimage.2019.116266.

466 **Aisen PS**, Siemers E, Michelson D, Salloway S, Sampaio C, Carrillo MC, Sperling R, Doody R, Scheltens P, Bateman
467 R, Weiner M, Vellas B. What Have We Learned from Expedition III and EPOCH Trials? Perspective of the CTAD
468 Task Force. J Prev Alzheimers Dis. 2018; 5(3):171–174.

469 **Akaike H**. In: Information Theory and an Extension of the Maximum Likelihood Principle New York, NY: Springer
470 New York; 1998. p. 199–213. https://doi.org/10.1007/978-1-4612-1694-0_15, doi: 10.1007/978-1-4612-1694-
471 0_15.

472 **Antelmi L**, Ayache N, Robert P, Lorenzi M. Sparse Multi-Channel Variational Autoencoder for the Joint Analysis
473 of Heterogeneous Data. In: *ICML 2019 - 36th International Conference on Machine Learning* Long Beach, United
474 States; 2019. .

475 **Bateman RJ**, Xiong C, Benzinger TLS, Fagan AM, Goate A, Fox NC, Marcus DS, Cairns NJ, Xie X, Blazey TM,
476 Holtzman DM, Santacruz A, Buckles V, Oliver A, Moulder K, Aisen PS, Ghetti B, Klunk WE, McDade E, Martins
477 RN, et al. Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease. New England Journal
478 of Medicine. 2012; 367(9):795–804. PMID: 22784036.

479 **Bilgel M**, Jedynak B, Wong DF, Resnick SM, Prince JL. Temporal Trajectory and Progression Score Estimation
480 from Voxelwise Longitudinal Imaging Measures: Application to Amyloid Imaging. Inf Process Med Imaging.
481 2015; 24:424–436.

482 **Blennow K**, Hampel H, Weiner M, Zetterberg H. Cerebrospinal fluid and plasma biomarkers in Alzheimer
483 disease. Nat Rev Neurol. 2010 Mar; 6(3):131–144.

484 **Braak H**, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 1991; 82(4):239–
485 259.

486 **Cash DM**, Frost C, Iheme LO, ?nay D, Kandemir M, Fripp J, Salvado O, Bourgeat P, Reuter M, Fischl B, Lorenzi
487 M, Frisoni GB, Pennec X, Pierson RK, Gunter JL, Senjem ML, Jack CR, Guizard N, Fonov VS, Collins DL, et al.
488 Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge.
489 Neuroimage. 2015 Dec; 123:149–164.

490 **Chen TQ**, Rubanova Y, Bettencourt J, Duvenaud DK. Neural Ordinary Differential Equations. In: Ben-
491 gio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, editors. *Advances in Neural Infor-*
492 *mation Processing Systems 31* Curran Associates, Inc.; 2018.p. 6571–6583. http://papers.nips.cc/paper/
493 7892-neural-ordinary-differential-equations.pdf.

494 **Cummings J**, Blennow K, Johnson K, Keeley M, Bateman RJ, Molinuevo JL, Touchon J, Aisen P, Vellas B. Anti-Tau
495 Trials for Alzheimer's Disease: A Report from the EU/US/CTAD Task Force. J Prev Alzheimers Dis. 2019;
496 6(3):157–163.

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

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

502 **Desikan RS**, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman
503 BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI
504 scans into gyral based regions of interest. NeuroImage. 2006 Jul; 31(3):968–980.

505 **Donohue MC**, Jacqmin-Gadda H, Goff ML, Thomas RG, Raman R, Gamst AC, Beckett LA, Jack CR, Weiner MW,
506 Dartigues JF, Aisen PS. Estimating long-term multivariate progression from short-term data. Alzheimer's &
507 Dementia. 2014; 10(5, Supplement):S400 – S410. doi: https://doi.org/10.1016/j.jalz.2013.10.003.

508 **Egan MF**, Kost J, Voss T, Mukai Y, Aisen PS, Cummings JL, Tariot PN, Vellas B, van Dyck CH, Boada M, Zhang Y, Li
509 W, Furtek C, Mahoney E, Harper Mozley L, Mo Y, Sur C, Michelson D. Randomized Trial of Verubecestat for
510 Prodromal Alzheimer's Disease. N Engl J Med. 2019 Apr; 380(15):1408–1420.

511 **Fonteijn HM**, Modat M, Clarkson MJ, Barnes J, Lehmann M, Hobbs NZ, Scahill RI, Tabrizi SJ, Ourselin S, Fox
512 NC, Alexander DC. An event-based model for disease progression and its application in familial Alzheimer's
513 disease and Huntington's disease. NeuroImage. 2012 Apr; 60(3):1880–1889.

514 **Gamberger D**, Lavrač N, Srivatsa S, Tanzi RE, Doraiswamy PM. Identification of clusters of rapid and slow
515 decliners among subjects at risk for Alzheimer's disease. Sci Rep. 2017 07; 7(1):6763.

516 **Garbarino S**, Lorenzi M. Modeling and Inference of Spatio-Temporal Protein Dynamics Across Brain Networks.
517 In: *IPMI 2019 - 26th International Conference on Information Processing in Medical Imaging*, vol. 11492 of LNCS
518 Hong-Kong, China: Springer; 2019. p. 57–69. https://hal.inria.fr/hal-02165021.

519 **Gauthier S**, Alam J, Fillit H, Iwatsubo T, Liu-Seifert H, Sabbagh M, Salloway S, Sampaio C, Sims JR, Sperling B,
520 Sperling R, Welsh-Bohmer KA, Touchon J, Vellas B, Aisen P. Combination Therapy for Alzheimer's Disease:
521 Perspectives of the EU/US CTAD Task Force. J Prev Alzheimers Dis. 2019; 6(3):164–168.

522 **Ghahramani Z**, Beal MJ. Graphical models and variational methods. In: Opper M, Saad D, editors. *Advanced*
523 *mean field methods: theory and practice* Neural Information Processing, MIT; 2001.

524 **Hao W**, Friedman A. Mathematical model on Alzheimer's disease. BMC Syst Biol. 2016 11; 10(1):108.

525 **Henley D**, Raghavan N, Sperling R, Aisen P, Raman R, Romano G. Preliminary Results of a Trial of Atabecestat in
526 Preclinical Alzheimer's Disease. N Engl J Med. 2019 04; 380(15):1483–1485.

527 **Honig LS**, Vellas B, Woodward M, Boada M, Bullock R, Borrie M, Hager K, Andreasen N, Scarpini E, Liu-Seifert
528 H, Case M, Dean RA, Hake A, Sundell K, Poole Hoffmann V, Carlson C, Khanna R, Mintun M, DeMattos R,
529 Selzler KJ, et al. Trial of Solanezumab for Mild Dementia Due to Alzheimer's Disease. N Engl J Med. 2018 01;
530 378(4):321–330.

531 **Howard R**, Liu KY. Questions EMERGE as Biogen claims aducanumab turnaround. Nat Rev Neurol. 2020 02;
532 16(2):63–64.

533 **Iturria-Medina Y**, Sotero RC, Toussaint PJ, Mateos-P?rez JM, Evans AC, Initiative ADN. Early role of vascular
534 dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. Nat Commun.
535 2016 06; 7:11934.

536 **Iturria-Medina Y**, Carbonell FM, Sotero RC, Chouinard-Decorte F, Evans AC. Multifactorial causal model
537 of brain (dis)organization and therapeutic intervention: Application to Alzheimer's disease. Neu-
538 roImage. 2017; 152:60 – 77. http://www.sciencedirect.com/science/article/pii/S1053811917301684, doi:
539 https://doi.org/10.1016/j.neuroimage.2017.02.058.

540 **Jack CR**, Holtzman DM. Biomarker modeling of Alzheimer's disease. Neuron. 2013 Dec; 80(6):1347–1358.

541 **Jack CR**, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD,
542 Lesnick TG, Pankratz VS, Donohue MC, Trojanowski JQ. Tracking pathophysiological processes in Alzheimer's
543 disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol. 2013 Feb; 12(2):207–216.

544 **Jedynak BM**, Lang A, Liu B, Katz E, Zhang Y, Wyman BT, Raunig D, Jedynak CP, Caffo B, Prince JL. A computational
545 neurodegenerative disease progression score: method and results with the Alzheimer's disease Neuroimaging
546 Initiative cohort. NeuroImage. 2012 Nov; 63(3):1478–1486.

547 **Kaufman SK**, Del Tredici K, Thomas TL, Braak H, Diamond MI. Tau seeding activity begins in the transen-
548 torhinal/entorhinal regions and anticipates phospho-tau pathology in Alzheimer's disease and PART. Acta
549 Neuropathologica. 2018 Jul; 136(1):57–67. https://doi.org/10.1007/s00401-018-1855-6, doi: 10.1007/s00401-
550 018-1855-6.

551 **Kim J**, Basak JM, Holtzman DM. The role of apolipoprotein E in Alzheimer's disease. Neuron. 2009 Aug;
552 63(3):287–303.

553 **Kingma DP**, Ba J, Adam: A Method for Stochastic Optimization; 2014. http://arxiv.org/abs/1412.6980, cite
554 arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning
555 Representations, San Diego, 2015.

556 **Kingma DP**, Welling M. Auto-Encoding Variational Bayes. CoRR. 2013; abs/1312.6114.

557 **Klein G**, Delmar P, Voyle N, Rehal S, Hofmann C, Abi-Saab D, Andjelkovic M, Ristic S, Wang G, Bateman R, Kerchner
558 GA, Baudler M, Fontoura P, Doody R. Gantenerumab reduces amyloid-β plaques in patients with prodromal
559 to moderate Alzheimer's disease: a PET substudy interim analysis. Alzheimer's Research & Therapy. 2019
560 Dec; 11(1):101. https://doi.org/10.1186/s13195-019-0559-z, doi: 10.1186/s13195-019-0559-z.

561 **Kochhann R**, Varela JS, Lisboa CSM, Chaves MLF. The Mini Mental State Examination: Review of cutoff points
562 adjusted for schooling in a large Southern Brazilian sample. Dement Neuropsychol. 2010; 4(1):35–41.

563 **Li D**, Iddi S, Thompson WK, Donohue MC. Bayesian latent time joint mixed effect models for multicohort
564 longitudinal data. Stat Methods Med Res. 2019 03; 28(3):835–845.

565 **Lorenzi M**, Filippone M, Frisoni GB, Alexander DC, Ourselin S. Probabilistic disease progression modeling to
566 characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer's disease. NeuroImage.
567 2017; doi: https://doi.org/10.1016/j.Neuroimage.2017.08.059.

Manuscript submitted to eLife

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

572 **Marinescu RV**, Eshaghi A, Lorenzi M, Young AL, Oxtoby NP, Garbarino S, Crutch SJ, Alexander DC. DIVE: A
573 spatiotemporal progression model of brain pathology in neurodegenerative disorders. NeuroImage. 2019 05;
574 192:166–177.

575 **Marinescu R**, Eshaghi A, Alexander D, Golland P. BrainPainter: A software for the visualisation of brain structures,
576 biomarkers and associated pathological processes. arXiv preprint arXiv:190508627. 2019; .

577 **Murphy MP**, LeVine H. Alzheimer's disease and the amyloid-beta peptide. J Alzheimers Dis. 2010; 19(1):311–323.

578 **Oxtoby NP**, Garbarino S, Firth NC, Warren JD, Schott JM, Alexander DC. Data-Driven Sequence of Changes to
579 Anatomical Brain Connectivity in Sporadic Alzheimer's Disease. Front Neurol. 2017; 8:580.

580 **Oxtoby NP**, Young AL, Cash DM, Benzinger TLS, Fagan AM, Morris JC, Bateman RJ, Fox NC, Schott JM, Alexander
581 DC. Data-driven models of dominantly-inherited Alzheimer's disease progression. Brain. 2018 05; 141(5):1529–
582 1544.

583 **Paszke A**, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A. Automatic
584 differentiation in PyTorch. . 2017; .

585 **Petrella JR**, Hao W, Rao A, Doraiswamy PM. Computational Causal Modeling of the Dynamic Biomarker Cascade
586 in Alzheimer's Disease. Comput Math Methods Med. 2019; 2019:6216530.

587 **Pontecorvo MJ**, Devous MD, Kennedy I, Navitsky M, Lu M, Galante N, Salloway S, Doraiswamy PM, Southekal S,
588 Arora AK, McGeehan A, Lim NC, Xiong H, Truocchio SP, Joshi AD, Shcherbinin S, Teske B, Fleisher AS, Mintun
589 MA. A multicentre longitudinal study of flortaucipir (18F) in normal ageing, mild cognitive impairment and
590 Alzheimer's disease dementia. Brain. 2019 06; 142(6):1723–1735.

591 **Prince M**, Wimo A, Guerchet M, Ali G, Wu YT, Prina M. World Alzheimer Report 2015 - The Global Impact of
592 Dementia: An analysis of prevalence, incidence, cost and trends. Alzheimer's Disease International; 2015.

593 **Reuter M**, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal
594 image analysis. NeuroImage. 2012 Jul; 61(4):1402–1418.

595 **Rowe CC**, Ellis KA, Rimajova M, Bourgeat P, Pike KE, Jones G, Fripp J, Tochon-Danguy H, Morandeau L, O'Keefe
596 G, Price R, Raniga P, Robins P, Acosta O, Lenzo N, Szoeke C, Salvado O, Head R, Martins R, Masters CL, et al.
597 Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol
598 Aging. 2010 Aug; 31(8):1275–1283.

599 **Safieh M**, Korczyn AD, Michaelson DM. ApoE4: an emerging therapeutic target for Alzheimer's disease. BMC
600 Medicine. 2019; 17.

601 **Schiratti J**, Allassonnière S, Colliot O, Durrleman S. Learning spatiotemporal trajectories from manifold-valued
602 longitudinal data. In: NIPS; 2015. p. 2404–2412.

603 **Schuff N**, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, Thompson PM, Jack CR, Weiner MW. MRI
604 of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain.
605 2009 Apr; 132(Pt 4):1067–1077.

606 **Shaw LM**, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A,
607 Lewczuk P, Dean R, Siemers E, Potter W, Lee VM, Trojanowski JQ. Cerebrospinal fluid biomarker signature in
608 Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009 Apr; 65(4):403–413.

609 **Sivera R**, Capet N, Manera V, Fabre R, Lorenzi M, Delingette H, Pennec X, Ayache N, Robert P. Voxel-based
610 assessments of treatment effects on longitudinal brain changes in the Multidomain Alzheimer Preventive
611 Trial cohort. Neurobiology of Aging. 2020; 94:50 – 59. http://www.sciencedirect.com/science/article/pii/
612 S0197458020301317, doi: https://doi.org/10.1016/j.neurobiolaging.2019.11.020.

613 **Sperling RA**, Jack CR, Aisen PS. Testing the right target and right drug at the right stage. Sci Transl Med. 2011
614 Nov; 3(111):111cm33.

615 **Villemagne VL**, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, Szoeke C, Macaulay SL, Martins R, Maruff
616   P, Ames D, Rowe CC, Masters CL. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic
617   Alzheimer's disease: a prospective cohort study. Lancet Neurol. 2013 Apr; 12(4):357–367.

618 **Wessels AM**, Tariot PN, Zimmer JA, Selzler KJ, Bragg SM, Andersen SW, Landry J, Krull JH, Downing AM, Willis BA,
619   Shcherbinin S, Mullen J, Barker P, Schumi J, Shering C, Matthews BR, Stern RA, Vellas B, Cohen S, MacSweeney E,
620   et al. Efficacy and Safety of Lanabecestat for Treatment of Early and Mild Alzheimer Disease: The AMARANTH
621   and DAYBREAK-ALZ Randomized Clinical Trials. JAMA Neurol. 2019 Nov; .

622 **Westwood S**, Leoni E, Hye A, Lynham S, Khondoker MR, Ashton NJ, Kiddle SJ, Baird AL, Sainz-Fuertes R, Leung
623   R, Graf J, Hehir CT, Baker D, Cereda C, Bazenet C, Ward M, Thambisetty M, Lovestone S. Blood-Based
624   Biomarker Candidates of Cerebral Amyloid Using PiB PET in Non-Demented Elderly. J Alzheimers Dis. 2016
625   03; 52(2):561–572.

626 **Young AL**, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, Schott JM, Alexander DC. A data-driven model of
627   biomarker changes in sporadic Alzheimer's disease. Brain. 2014 Sep; 137(Pt 9):2564–2577.

628 **Zetterberg H**, Burnham SC. Blood-based molecular biomarkers for Alzheimer's disease. Molecular brain. 2019
629   Mar; 12(1):26–26. doi: 10.1186/s13041-019-0448-1.

Manuscript submitted to eLife
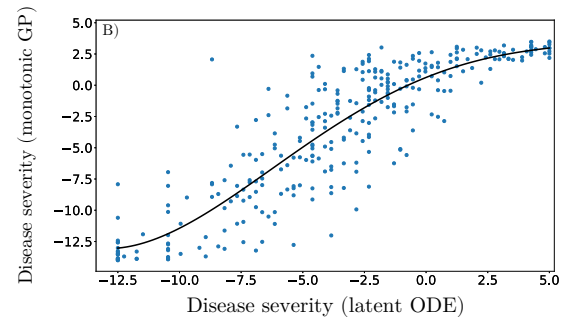
Appendix 1

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 (*Lorenzi et al., 2017*) 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.

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.
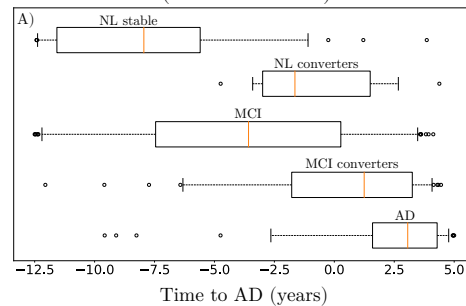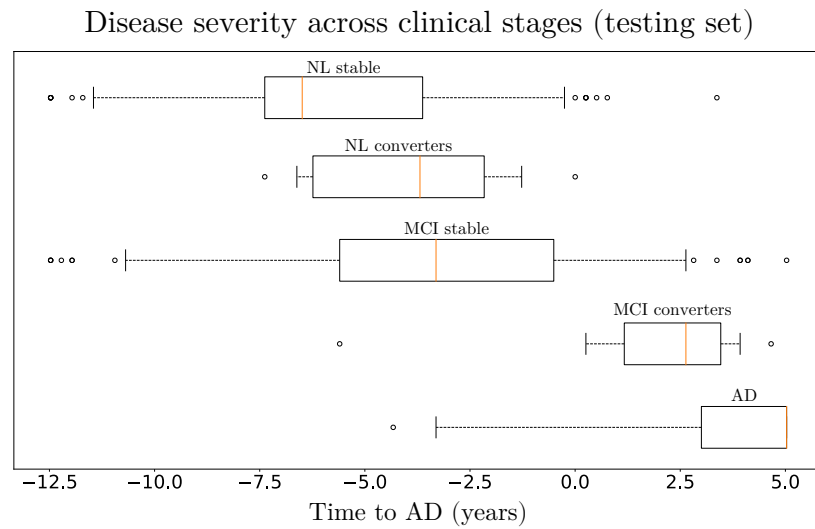

Appendix 1 Figure 1. A: Distribution of the disease severity estimated by the monotonic GP method (*Lorenzi et al., 2017*) on the training set. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer's dementia. B: Comparison of the disease severity estimated by our method (denoted as latent ODE) with respect to the one estimated by the monotonic GP.

Appendix 1 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 (cm³)	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)

Appendix 1 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.

Appendix 2

Simulated clinical endpoints

We provide in *Table 1* the estimated values for each clinical score at predicted conversion time for the normal progression case when performing the simulations presented in Section Simulating clinical intervention.

Appendix 2 Table 1. 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)

Appendix 3

Lower bound

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

$$
\begin{aligned}
\log p(\mathbf{X}|\sigma^2, \boldsymbol{\psi}) &= \log\left[\int p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \sigma^2, \boldsymbol{\psi})p(\mathbf{z})d\mathbf{z}\right] \\
&= \log\left[\int p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \sigma^2, \boldsymbol{\psi})p(\mathbf{z})\frac{q(\mathbf{z}|\mathbf{X})}{q(\mathbf{z}|\mathbf{X})}d\mathbf{z}\right] \\
&= \log\left[\mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\frac{p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \sigma^2, \boldsymbol{\psi})p(\mathbf{z})}{q(\mathbf{z}|\mathbf{X})}\right] \\
&\overset{\text{Jensen}}{\geq} \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\left[\log\frac{p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \sigma^2, \boldsymbol{\psi})p(\mathbf{z})}{q(\mathbf{z}|\mathbf{X})}\right] \\
&= \mathbb{E}_{q(\mathbf{z}|\mathbf{X})}\left[\log p(\mathbf{X}|\mathbf{z}, \theta_{ODE}, \sigma^2, \boldsymbol{\psi})\right] - \mathcal{D}\left[q(\mathbf{z}|\mathbf{X})|p(\mathbf{z})\right] \\
&= \mathcal{E}.
\end{aligned}
\tag{10}
$$

Given that:

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

We obtain:

$$
\mathcal{E} = \sum_m E_{q(\mathbf{z}|\mathbf{X})}\left[\log p(\mathbf{x}^m|\mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m)\right] - \mathcal{D}\left[q(z^m|\mathbf{x}^m)|p(z^m)\right].
\tag{11}
$$

KL divergence

We have that:

$$
\begin{aligned}
q(z^m|\mathbf{X}) &= \mathcal{N}(f(\mathbf{x}^m, \phi_m^1), h(\mathbf{x}^m, \phi_m^2)), \\
p(z^m) &= \mathcal{N}(0, 1).
\end{aligned}
\tag{12}
$$

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

$$
\begin{aligned}
\mathcal{D}\left[q(\mathbf{z}|\mathbf{X})|p(\mathbf{z})\right] &= \sum_m \mathcal{D}\left[q(z^m|\mathbf{x}^m)|p(z^m)\right] \\
&= \frac{1}{2}\sum_m\left[-\log(h(\mathbf{x}^m, \phi_m^2)) - 1 + h(\mathbf{x}^m, \phi_m^2) + f(\mathbf{x}^m, \phi_m^1)^2\right]
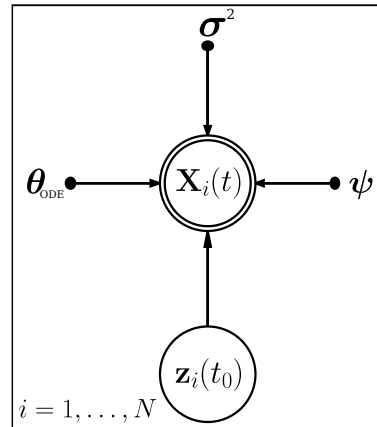\end{aligned}
\tag{13}
$$

Graphical model

Figure 1 below provides the graphical model illustrating the method presented in Section Methods.

ELBO computation

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

Appendix 3 Figure 1. Graphical model of the proposed method.

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

1: **function** COMPUTE_ELBO($\mathbf{X}(t), \mathbf{X}(t_0), \theta_{ODE}, \psi, \phi, \sigma^2$)
   For ease of notation we drop the i index in the pseudo-code.
2:   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).
3:   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.
4:   Compute $\mathbb{E}_{q(\mathbf{z}(t_0)|\mathbf{X}(t_0))}\left[\log p(\mathbf{x}^m(t)|\mathbf{z}(t), \theta_{ODE}, \sigma_m^2, \psi_m)\right] \approx -\frac{D_m}{2}\log(2\pi\sigma_m^2) - \frac{1}{2\sigma_m^2}||\mathbf{x}^m(t) - \mu_m(\mathbf{z}(t))||^2$ ▷ Expectation term *Equation 6*.
5:   Compute $\mathcal{D}\left[q(z^m(t_0)|\mathbf{x}^m(t_0))|p(z^m(t_0))\right] = \frac{1}{2}\left[-\log(h(\mathbf{x}^m(t_0), \phi_m^2)) - 1 + h(\mathbf{x}^m(t_0), \phi_m^2) + f(\mathbf{x}^m(t_0), \phi_m^1)^2\right]$ ▷ KL divergence *Equation 6*.
6:   Compute $\mathcal{E} = \sum_m \mathbb{E}_{q(\mathbf{z}(t_0)|\mathbf{X}(t_0))}\left[\log p(\mathbf{x}^m(t)|\mathbf{z}(t), \theta_{ODE}, \sigma_m^2, \psi_m)\right] - \mathcal{D}\left[q(z^m(t_0)|\mathbf{x}^m(t_0))|p(z^m(t_0))\right]$.
7:   **Return** $\mathcal{E}$
8: **end function**