



HAL
open science

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

Clément Abi Nader, Nicholas Ayache, Giovanni B Frisoni, Philippe Robert,
Marco Lorenzi

► To cite this version:

Clément Abi Nader, Nicholas Ayache, Giovanni B Frisoni, Philippe Robert, Marco Lorenzi. Simulating the outcome of amyloid treatments in Alzheimer's Disease from multi-modal imaging and clinical data. 2020. hal-02968724v1

HAL Id: hal-02968724

<https://hal.science/hal-02968724v1>

Preprint submitted on 16 Oct 2020 (v1), last revised 10 Feb 2021 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Clément Abi Nader^{a, 1}, Nicholas Ayache^a, Giovanni B. Frisoni^b, Philippe Robert^c, and Marco Lorenzi^a, for the Alzheimer's Disease Neuroimaging Initiative²

^aUniversité Côte d'Azur, Inria Sophia Antipolis, Epione Research Project, 2004 route des Lucioles, 06902 Sophia-Antipolis, France; ^bMemory Clinic and LANVIE-Laboratory of Neuroimaging of Aging, Hospitals and University of Geneva, Rue Gabrielle-Perret-Gentil 6, 1205 Genève, Switzerland; ^cUniversité Côte d'Azur, CoBTeK lab, MNC3 program, 10 rue Molière, 06100 Nice, France.

This manuscript was compiled on October 8, 2020

1 **Recent failures of clinical trials in Alzheimer's Disease underline the**
2 **critical importance of identifying optimal intervention time to maxi-**
3 **mize cognitive benefit. While several models of disease progression**
4 **have been proposed, we still lack quantitative approaches simulat-**
5 **ing the effect of treatment strategies on the clinical evolution. In this**
6 **work, we present a data-driven method to model dynamical relation-**
7 **ships between imaging and clinical biomarkers. Our approach allows**
8 **simulating intervention at any stage of the pathology by modulat-**
9 **ing the progression speed of the biomarkers, and by subsequently**
10 **assessing the impact on disease evolution. When applied to multi-**
11 **modal imaging and clinical data from the Alzheimer's Disease Neu-**
12 **roimaging Initiative our method enables to generate hypothetical sce-**
13 **narios of amyloid lowering interventions. Our results show that in a**
14 **study with 1000 individuals per arm, accumulation should be com-**
15 **pletely arrested at least 5 years before Alzheimer's dementia diagno-**
16 **sis to lead to statistically powered improvement of clinical endpoints.**

Alzheimer's Disease | Clinical trials | Dynamical systems | Variational inference

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

29 about the critical importance of intervention time, and about
30 the need of starting anti-amyloid treatments during the
31 pre-symptomatic stages of the disease (15). Nevertheless,
32 the design of optimal intervention strategies is currently not
33 supported by quantitative analysis methods allowing to model
34 and assess the effect of intervention time and dosing (16). The
35 availability of models of the pathophysiology of AD would
36 entail great potential to test and analyze clinical hypothesis
37 characterizing AD mechanisms, progression, and intervention
38 scenarios.

40 Within this context, quantitative models of disease progression,
41 referred to as Disease Progression Models (DPMs), have been
42 proposed (17–21), to quantify the dynamics of the changes
43 affecting the brain during the whole disease span. These
44 models rely on the statistical analysis of large datasets of
45 different data modalities, such as clinical scores, or brain
46 imaging measures derived from Magnetic Resonance Imaging
47 (MRI), Amyloid- and Fluorodeoxyglucose-Positron Emission
48 Tomography (PET) (22–24). In general, DPMs estimate

Significance Statement

This study investigates a novel quantitative instrument for the development of intervention strategies for disease modifying drugs in Alzheimer's Disease. Our framework enables the simulation of the effect of intervention time and drug dosage on the evolution of imaging and clinical biomarkers in clinical trials. The proposed data-driven approach is based on the modeling of the spatio-temporal dynamics governing the joint evolution of imaging and clinical measurements throughout the disease. Our experimental simulations are compatible with the outcomes observed in past clinical trials, and suggest that anti-amyloid treatments should be administered at least 5 years earlier than what is currently being done in order to obtain statistically powered improvement of clinical endpoints.

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 B. Frisoni, and Philippe Robert.

The authors declare no competing interests.

¹To whom correspondence should be addressed. E-mail: clement.abi-nader@inria.fr

²Data 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.

49 a long-term disease evolution from the joint analysis of
50 multivariate time-series acquired on a short-term time-scale.
51 Due to the temporal delay between the disease onset and
52 the appearance of the first symptoms, DPMs rely on the
53 identification of an appropriate temporal reference to describe
54 the long-term disease evolution (25, 26). These tools are
55 promising approaches for the analysis of clinical trials data, as
56 they allow to represent the longitudinal evolution of multiple
57 biomarkers through a global model of disease progression.
58 Such a model can be subsequently used as a reference in
59 order to stage subjects and quantify their relative progression
60 speed (27–29). However, these approaches remain purely
61 descriptive as they don’t account for causal relationships
62 among biomarkers. Therefore, they generally don’t allow
63 to simulate progression scenarios based on hypothetical
64 intervention strategies, thus providing a limited interpretation
65 of the pathological dynamics. This latter capability is of
66 utmost importance for planning and assessment of disease
67 modifying treatments.

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

92 In this work we present a novel computational model of AD
93 progression allowing to simulate intervention strategies across
94 the history of the disease. The model is here used to quantify
95 the potential effect of amyloid modifiers on the progression
96 of brain atrophy, glucose hypometabolism, and ultimately on
97 the clinical outcomes for different scenarios of intervention.
98 To this end, we model the joint spatio-temporal variation
99 of different modalities along the history of AD by identifying
100 a system of ODEs governing the pathological progression.
101 This latent ODEs system is specified within an interpretable
102 low-dimensional space relating multi-modal information, and
103 combines clinically-inspired constraints with unknown inter-
104 actions that we wish to estimate. The interpretability of the
105 relationships in the latent space is ensured by mapping each
106 data modality to a specific latent coordinate. The model
107 is formulated within a Bayesian framework, where the latent
108 representation and dynamics are efficiently estimated
109 through stochastic variational inference. To generate hypo-

110 thetical scenarios of amyloid lowering interventions, we apply
111 our approach to multi-modal imaging and clinical data from
112 the Alzheimer’s Disease Neuroimaging Initiative (ADNI). Our
113 results provide a meaningful quantification of different interven-
114 tion strategies, compatible with findings previously reported in
115 clinical studies. For example, we estimate that in a study with
116 100 individuals per arm, statistically powered improvement
117 of clinical endpoints can be obtained by completely arresting
118 amyloid accumulation at least 8 years before Alzheimer’s de-
119 mentia. The minimum intervention time decreases to 5 years
120 for studies based on 1000 individuals per arm.

121 Results

122 In the following sections, healthy individuals will be denoted
123 as NL stable, subjects with mild cognitive impairment as
124 MCI stable, subjects diagnosed with Alzheimer’s dementia as
125 AD, subjects progressing from NL to MCI as NL converters,
126 and subjects progressing from MCI to AD as MCI converters.
127 Amyloid concentration and glucose metabolism are respec-
128 tively measured by (18)F-florbetapir Amyloid (AV45)-PET
129 and (18)F-fluorodeoxyglucose (FDG)-PET imaging. Cognitive
130 and functional abilities are assessed by the following neuro-
131 psychological tests: Alzheimer’s Disease Assessment Scale
132 (ADAS11), Mini-Mental State Examination (MMSE), Func-
133 tional Assessment Questionnaire (FAQ), Rey Auditory Verbal
134 Learning Test (RAVLT) immediate, and RAVLT forgetting.

135 **Study cohort and biomarkers’ changes across clinical**
136 **groups.** Our study is based on a cohort of 311 amyloid
137 positive individuals composed of 46 NL stable subjects, 10
138 NL converters subjects, 106 subjects diagnosed with MCI,
139 76 MCI converters subjects, and 73 AD patients. The term
140 “amyloid positive” refers to subjects whose amyloid level in the
141 cerebrospinal fluid (CSF) was below the nominal cutoff of 192
142 pg/ml (35) either at baseline, or during any follow-up visit,
143 and conversion to AD was determined using the last available
144 follow-up information. The length of follow-up varies between
145 subjects and goes from 0 to 6 years. Further information about
146 the data are available on <https://adni.bitbucket.io/reference/>,
147 while details on data acquisition and processing are provided
148 in Section 1. We show in Table 1A socio-demographic
149 information for the training cohort across the different clinical
150 groups. Table 1B shows baseline values and annual rates of
151 change across clinical groups for amyloid burden (average
152 normalized AV45 uptake in frontal cortex, anterior cingulate,
153 precuneus and parietal cortex), glucose hypometabolism
154 (average normalized FDG uptake in frontal cortex, anterior
155 cingulate, precuneus and parietal cortex), for hippocampal
156 and medial temporal lobe volumes, and for the cognitive
157 ability as measured by ADAS11. Compatibly with previously
158 reported results (36, 37), we observe that while regional
159 atrophy, hypometabolism and cognition show increasing
160 rate of change when moving from healthy to pathological
161 conditions, the change of AV45 is maximum in NL stable and
162 MCI stable subjects. We also notice the increased magnitude
163 of ADAS11 in AD as compared to the other clinical groups.
164 Finally, the magnitude of change of FDG is generally milder
165 than the atrophy rates.

166 The observations presented in Table 1 provide us with a glimpse
167 into the biomarkers’ trajectories characterising AD. The com-
168

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]

plexity 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).

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 .

Progression model and latent relationships. We show in Figure 2 Panel I) 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

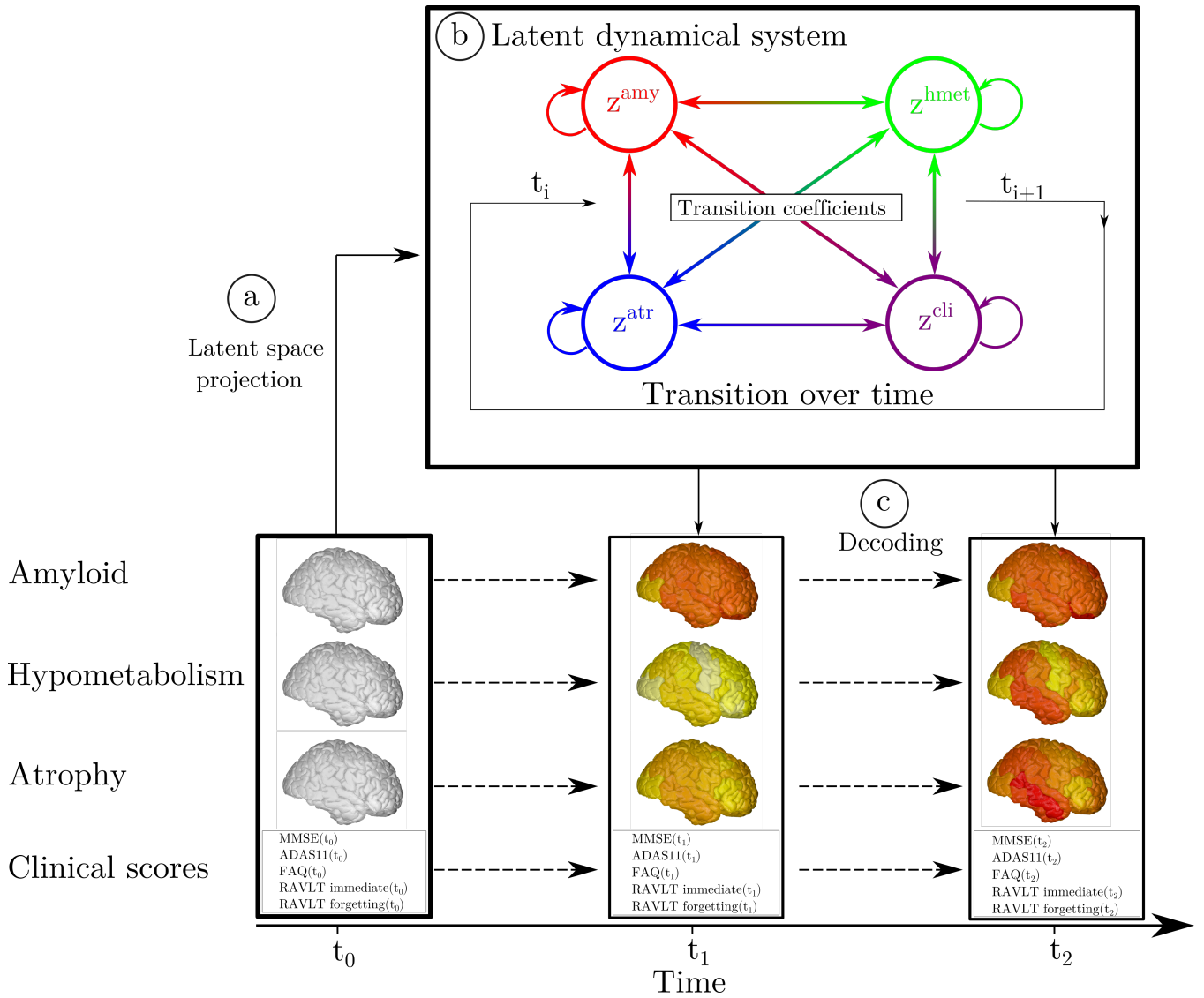


Fig. 1. Overview of the method. a) High-dimensional multi-modal measures are projected into a 4-dimensional latent space. Each data modality is transformed in a corresponding z-score z^{amy} , z^{hmet} , z^{atr} , z^{cli} . b) The dynamical system describing the relationships between the z-scores allows 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.

231 contribute to an increase of z^{cli} (Figure 2E). Finally, Figure
232 2F shows that high atrophy values lead to an increase mostly
233 along the clinical dimension z^{cli} . This chain of relation-
234 ships is in agreement with the cascade hypothesis of AD (2, 3).
235

236 Relying on the dynamical relationships shown in Figure
237 2 Panel I), starting from any initial set of biomarkers
238 values we can estimate the relative trajectories over time.
239 Figure 2 Panel II) (left), shows the evolution obtained by
240 extrapolating backward and forward in time the trajectory
241 associated to the z-scores of the AD group. The x-axis
242 represents the years from conversion to AD, where the
243 instant $t=0$ corresponds to the average time of diagnosis
244 estimated for the group of MCI progressing to dementia.
245 As observed in Figure 2 Panel I) and Table 1, the amyloid
246 score z^{amy} increases and saturates first, followed by z^{hmet}
247 and z^{atr} scores whose progression slows down when reaching
248 clinical conversion, while the clinical score exhibits strong
249 acceleration in the latest progression stages. Figure 2 Panel
250 II) (right) shows the group-wise distribution of the disease
251 severity estimated for each subject relatively to the modelled
252 long-term latent trajectories (Section Evaluating disease
253 severity). The group-wise difference of disease severity across
254 groups is statistically significant and increases when going
255 from healthy to pathological stages (Wilcoxon-Mann-Whitney
256 test $p < 0.001$ for each comparisons). The reliability of the
257 estimation of disease severity was further assessed through
258 testing on an independent cohort, and by comparison with
259 a previously proposed disease progression modeling method
260 from the state-of-the-art (25). The results are provided in
261 section 1 of the Supplementary Material and show positive
262 generalization results as well as a favourable comparison with
263 the benchmark method.
264

265 From the z-score trajectories of Figure 2 Panel II) (left) we
266 predict the progression of imaging and clinical measures shown
267 in Figure 3. We observe that amyloid load globally increases
268 and saturates early, compatibly with the positive amyloid
269 condition of the study cohort. Glucose hypometabolism and
270 gray matter atrophy increase are delayed with respect to
271 amyloid, and tend to map prevalently temporal and parietal
272 regions. Finally, the clinical measures exhibit a non-linear
273 pattern of change, accelerating during the latest progression
274 stages. These dynamics are compatible with the summary
275 measures on the raw data reported in Table 1. Finally, the
276 estimated progression of the RAVLT immediate and RAVLT
277 forgetting is provided in Figure S3 of the Supplementary
278 Material.
279

280 **Simulating clinical intervention.** This experimental section
281 is based on two intervention scenarios: a first one in which
282 amyloid is lowered by 100%, and a second one in which it
283 is reduced by 50% with respect to the estimated natural
284 progression. In Figure 4 we show the latent z-scores evolution
285 resulting from either 100% or 50% amyloid lowering performed
286 at the time $t = -12.5$ years. According to these scenarios,
287 intervention results in a sensitive reduction of the pathological
288 progression for atrophy, hypometabolism and clinical scores,
289 albeit with a stronger effect in case of total blockage.
290

291 We further estimated the resulting clinical endpoints associ-

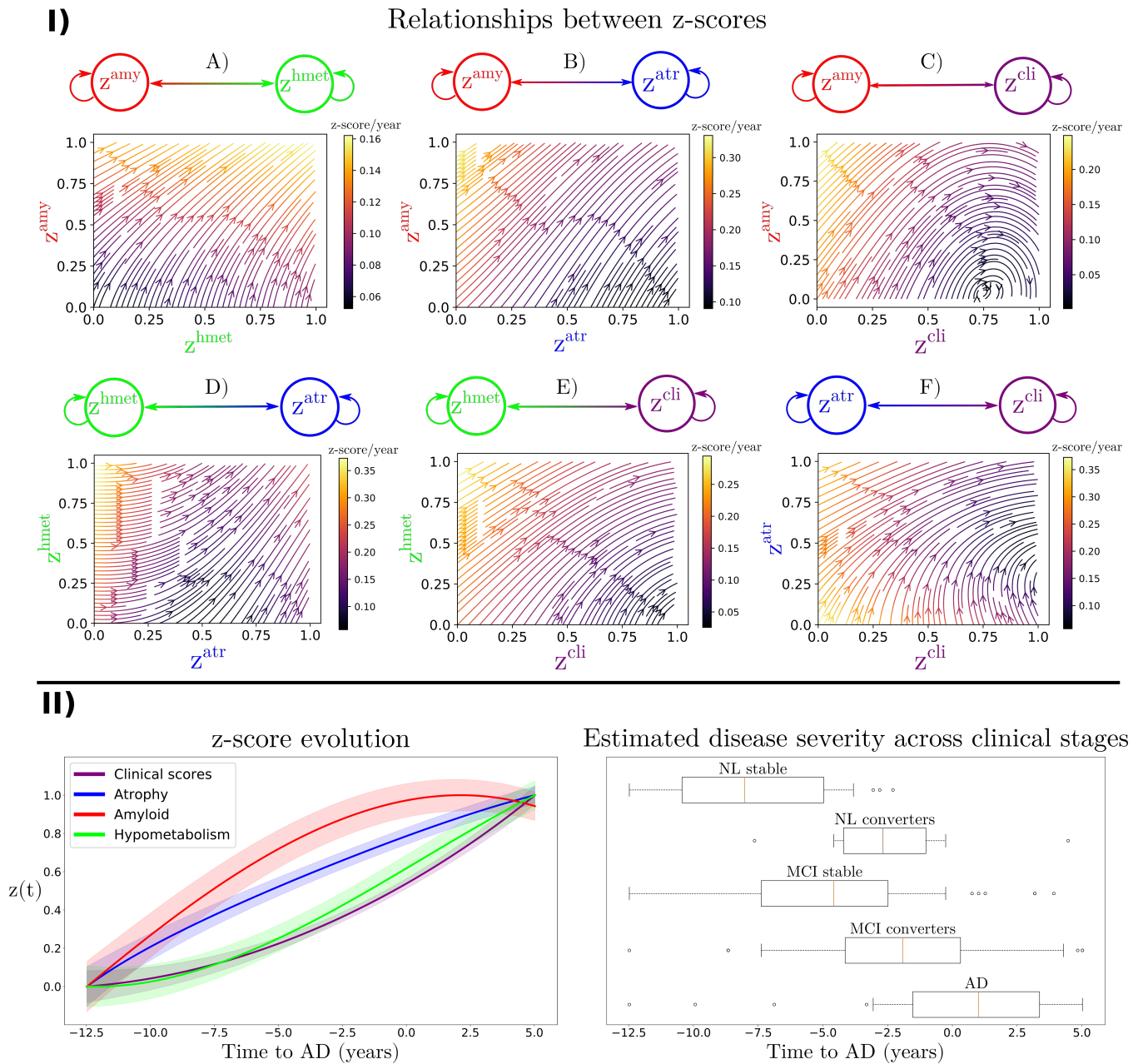
292 ated with the two amyloid lowering scenarios, at increasing
293 time points and for different sample sizes. Clinical endpoints
294 consisted in the simulated ADAS11, MMSE, and FAQ at
295 the reference conversion time ($t=0$). The case placebo
296 indicates the scenario where clinical values were computed
297 at conversion time from the estimated natural progression
298 shown in Figure 2 Panel II) (left). Figure 5 shows the change
299 in statistical power depending on intervention time and
300 sample sizes. For large sample sizes (1000 subjects per arm)
301 a power greater than 0.8 can be obtained around 5 years
302 before conversion, depending on the outcome score. When
303 sample size is lower than 100 subjects per arm, a power
304 greater than 0.8 is reached if intervention is performed at
305 the latest 8 years before conversion, with a mild variability
306 depending on the considered clinical score. We notice that
307 in the case of a 50% amyloid lowering, in order to reach the
308 same power intervention needs to be consistently performed
309 earlier compared to the scenario of 100% amyloid lowering
310 for the same sample size and clinical score. For instance, if
311 we consider ADAS11 with a sample size of 100 subjects per
312 arm, a power of 0.8 is obtained for a 100% amyloid lowering
313 intervention performed 8 years before conversion, while in
314 case of a 50% amyloid lowering the equivalent effect would be
315 obtained by intervening 10.5 years before conversion. Finally,
316 we also provide in Figure S4 of the Supplementary Material
317 the power analysis for the RAVLT immediate and RAVLT
318 forgetting.
319

320 We provide in Table 2 the estimated improvement for ADAS11,
321 MMSE, and FAQ at conversion with a sample size of 100
322 subjects per arm for both 100% and 50% amyloid lowering de-
323 pending on the intervention time. The estimated improvement
324 for RAVLT immediate and RAVLT forgetting is provided in
325 Table S2 in the Supplementary Material. We observe that for
326 the same intervention time, 100% amyloid lowering always
327 results in a larger improvement of clinical endpoints compared
328 to 50% amyloid lowering. We also note that in the case of
329 100% lowering, clinical endpoints obtained for intervention at
330 $t=-10$ years correspond to typical cutoff values for inclusion
331 into AD trials (ADAS11= 13.4 ± 6.2 , MMSE= 25.8 ± 2.5 , see
332 Table S3 in Supplementary Material) (39, 40).
333

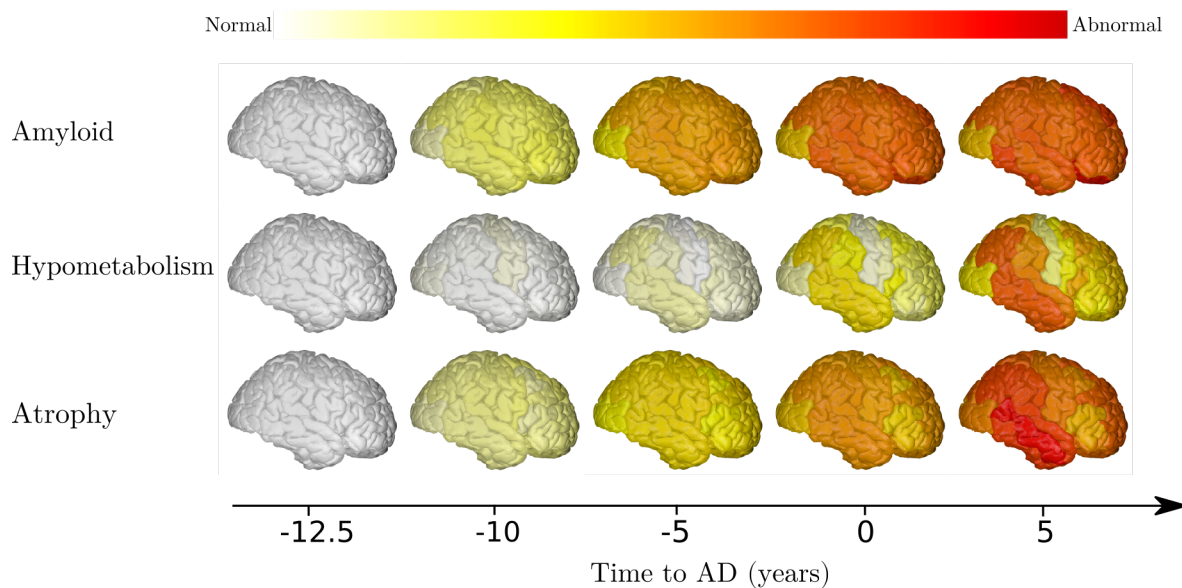
333 Discussion

334 We presented a framework to jointly model the progression
335 of multi-modal imaging and clinical data, based on the
336 estimation of latent biomarkers' relationships governing AD
337 progression. The model is designed to simulate intervention
338 scenarios in clinical trials, and in this study we focused on
339 assessing the effect of anti-amyloid drugs on biomarkers'
340 evolution, by quantifying the effect of intervention time and
341 drug efficacy on clinical outcomes. Our results underline the
342 critical importance of intervention time, which should be
343 performed sensibly early during the pathological history to
344 effectively appreciate the effectiveness of disease modifiers.
345

346 The results obtained with our model are compatible with
347 findings reported in recent clinical studies (11–13). For
348 example, if we consider 500 patients per arm and perform a
349 100% amyloid lowering intervention for 2 years to reproduce
350 the conditions of the recent trial of Verubecestat (12), the
351 average improvement of MMSE predicted by our model is of



Imaging evolution



Clinical outcomes evolution

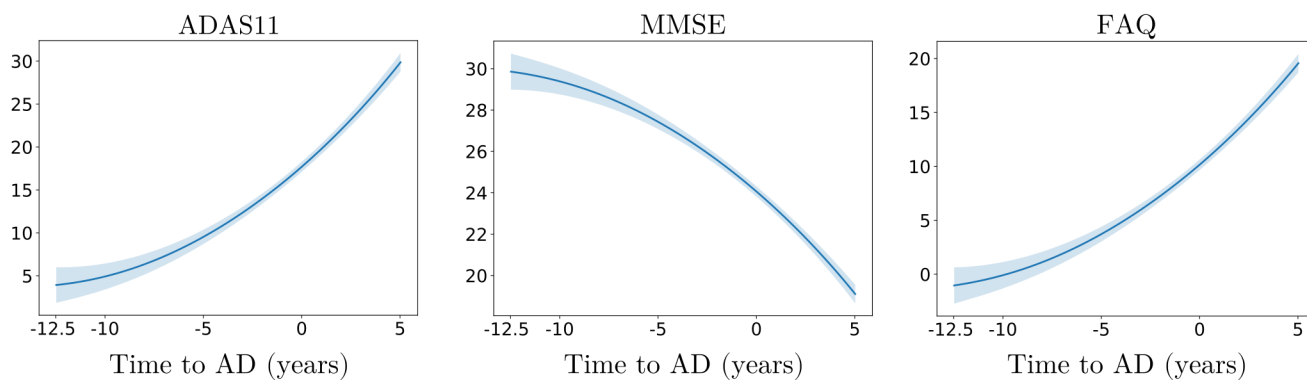


Fig. 3. Modelled long-term evolution of cortical measurements for the different types of imaging markers, and clinical scores. Shaded areas represent the standard deviation of the average trajectory. Brain images were generated using the software provided in (38).

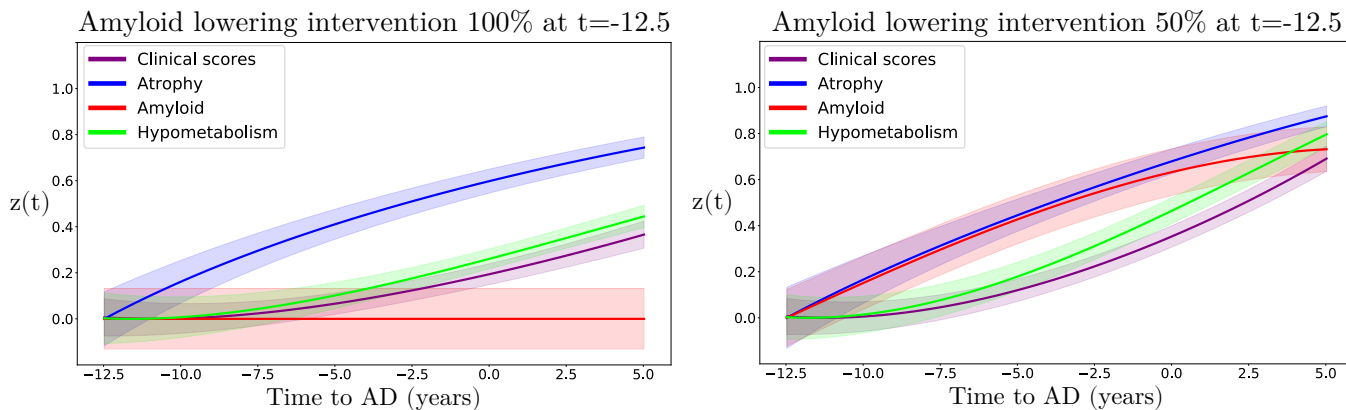


Fig. 4. Hypothetical scenarios of irreversible amyloid lowering interventions at t=-12.5 years from Alzheimer's dementia diagnosis, with a rate of 100 % (left) or 50 % (right). Shaded areas represent the standard deviation of the average trajectory.

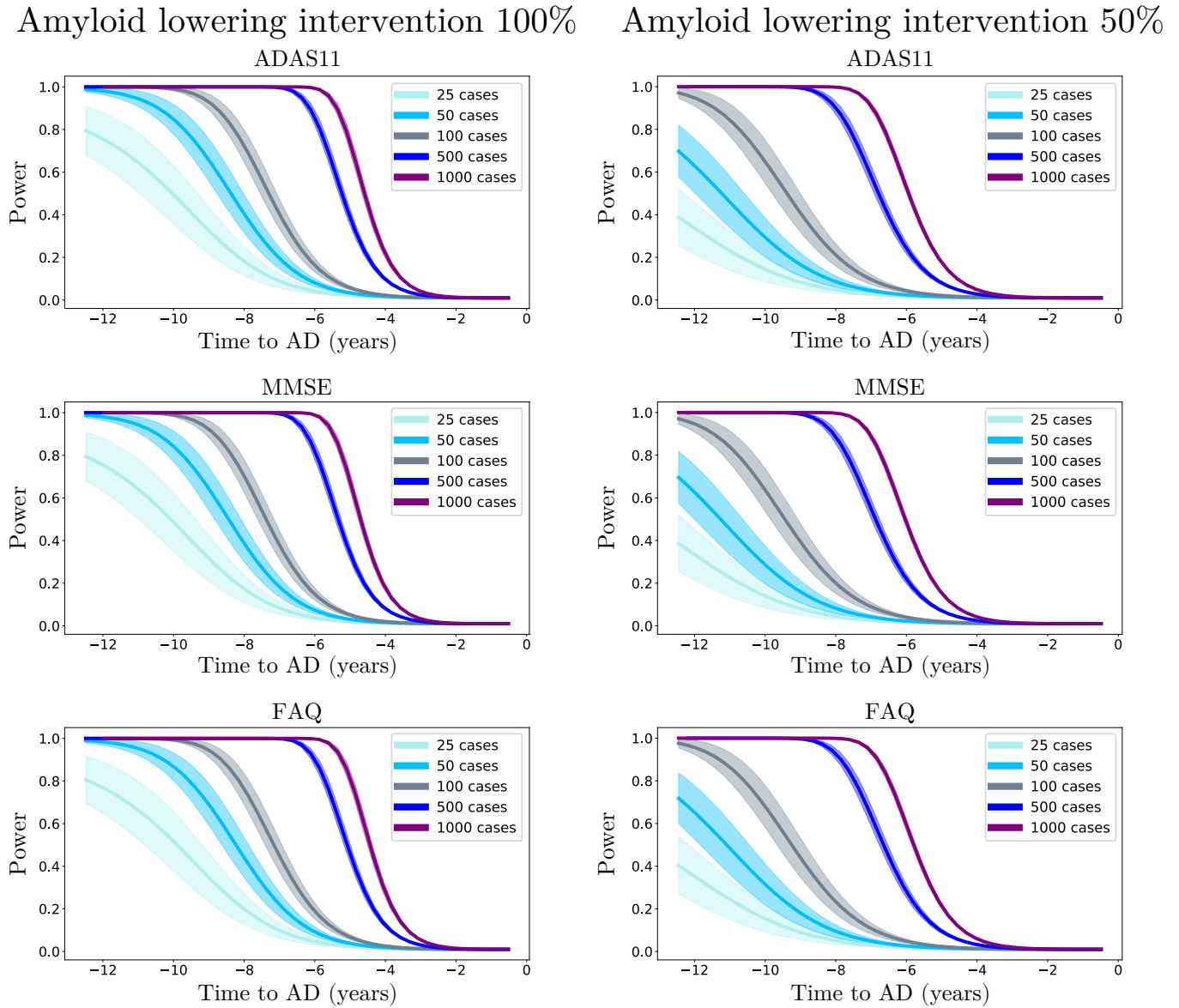


Fig. 5. Statistical power of the Student t-test comparing the estimated clinical outcomes at conversion time between placebo and treated scenarios, according to the year of simulated intervention (100% and 50% amyloid lowering) and sample size.

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.

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)

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)

0.02, falling in the 95% confidence interval measured during that study ($[-0.5 ; 0.8]$). While recent anti-amyloid trials such as (11–13) 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 5 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 2 Panel II) (right). 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 Panel I) 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 (51), 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 (52). 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 (53, 54). 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 3 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 4, in which we notice that atrophy is less affected by intervention, implying that its evolution is here importantly decorrelated from amyloid burden.

Materials and 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).

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 (55), 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 (56). 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 .

Data modelling. We consider observations $\mathbf{X}_i(t) = [\mathbf{x}_i^1(t), \mathbf{x}_i^2(t), \dots, \mathbf{x}_i^M(t)]^T$, 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 $\mathbf{z}_i(t)$:

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

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:

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

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:

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

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

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

521 For ease of notation, we drop the i index, and dependence on t
522 and t_0 is made implicit. Within a Bayesian framework, we wish
523 to maximize \mathcal{L} in order to obtain a posterior distribution for the
524 latent variable \mathbf{z} . Since derivation of this quantity is generally not
525 tractable, we resort to stochastic variational inference to tackle the
526 optimization problem. We assume a $\mathcal{N}(\mathbf{0}, \mathbf{I})$ prior for $p(\mathbf{z})$, and
527 introduce an approximate posterior distribution $q(\mathbf{z}|\mathbf{X})$ (57), in
528 order to derive a lower-bound \mathcal{E} for the marginal log-likelihood:

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

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:

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

529 Details about the ELBO derivation and the computation of the KL
530 divergence are given in the Supplementary Material. A graphical
531 model of the method is also provided in Figure S5, while Supplement-
532 ary Algorithm 1 details the steps to compute the ELBO.

533 **Model optimization.** Using the reparameterization trick (58), we can
534 efficiently sample from the posterior distribution $q(\mathbf{z}(t_0)|\mathbf{X}(t_0))$
535 to approximate the expectation terms. Moreover, thanks to our
536 choices of priors and approximations the KL terms can be computed
537 in closed-form. In practice, we sample from $q(\mathbf{z}(t_0)|\mathbf{X}(t_0))$ to
538 obtain a latent representation $\mathbf{z}(t_0)$ at baseline, while the follow-up
539 points are estimated by decoding the latent time-series obtained
540 through the integration of the ODEs of Eq 3. The model is
541 trained by computing the total ELBO for all the subjects at all
542 the available time points. The parameters $\psi, \phi^1, \phi^2, \theta_{ODE}, \sigma$
543 are optimized using gradient descent, which requires to backpropagate
544 through the integration operation.

545
546 In order to enable backpropagation through the ODEs integration
547 we need to numerically solve the differential equation using only
548 operations that can be differentiated. In this work, we used the
549 Midpoint method which follows a second order Runge-Kutta scheme.
550 The method consists in evaluating the derivative of the solution at
551 $(t_{i+1} + t_i)/2$, which is the midpoint between t_i at which the correct
552 $\mathbf{z}(t)$ is evaluated, and the following t_{i+1} :

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

553 Therefore, solving the system of Equation 3 on the interval
554 $[t_0, \dots, t]$ only requires operations that can be differentiated,
555 allowing to compute the derivatives of the ELBO with respect
556 to all the parameters, and to optimize them by gradient descent.
557 Moreover, in order to control the variability of the estimated
558 latent trajectory $\mathbf{z}(t)$ due to the error propagation during
559 integration, we initialized the weights of ϕ^1 and ϕ^2 such that
560 the approximate posterior of the latent representation for each
561 modality m at baseline was following a $\mathcal{N}(0, 0.01)$ distribution.
562 Finally, we also tested other ODE solvers such as Runge-Kutta
563 4, which gave similar results than the Midpoint method with a
564 slower execution time due its more expensive approximation scheme.
565

Concerning the implementation, we trained the model using the
ADAM optimizer (59) 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 (60), and we used the *torchdiffeq*
package developed in (61) 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
(62). 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 $\mathbf{Z}(t) = [\mathbf{z}_1(t), \dots, \mathbf{z}_N(t)]$ 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 1 each coordinate of the latent
representation can be interpreted as representing a single data
modality. Therefore, we propose to simulate the effect of an hypo-
thetical intervention on the disease progression, by modulating the
vector $\frac{dz(t)}{dt}$ after each integration step such that:

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

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. $\gamma_{amy} = 0.5$).

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:

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

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.

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

618 Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Health- 700
619 care; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & 701
620 Development, LLC.; Johnson & Johnson Pharmaceutical Research 702
621 & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso 703
622 Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technolo- 704
623 gies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal 705
624 Imaging; Servier; Takeda Pharmaceutical Company; and Transition 706
625 Therapeutics. The Canadian Institutes of Health Research is provid- 707
626 ing funds to support ADNI clinical sites in Canada. Private sector 708
627 contributions are facilitated by the Foundation for the National 709
628 Institutes of Health (www.fnih.org). The grantee organization is 710
629 the Northern California Institute for Research and Education, and 711
630 the study is coordinated by the Alzheimer's Therapeutic Research 712
631 Institute at the University of Southern California. ADNI data are 713
632 disseminated by the Laboratory for Neuro Imaging at the University 714
633 of Southern California. 715

- 634 1. M Prince, et al., *World Alzheimer Report 2015 - The Global Impact of Dementia: An analysis* 716
635 *of prevalence, incidence, cost and trends.* (Alzheimer's Disease International), (2015). 717
- 636 2. CR Jack, DM Holtzman, Biomarker modeling of Alzheimer's disease. *Neuron* **80**, 1347–1358 718
637 (2013). 719
- 638 3. CR Jack, et al., Tracking pathophysiological processes in Alzheimer's disease: an updated 720
639 hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**, 207–216 (2013). 721
- 640 4. VL Villemagne, et al., Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic 722
641 Alzheimer's disease: a prospective cohort study. *Lancet Neurol* **12**, 357–367 (2013). 723
- 642 5. MP Murphy, H LeVine, Alzheimer's disease and the amyloid-beta peptide. *J. Alzheimers Dis.* 724
643 **19**, 311–323 (2010). 725
- 644 6. A Delacourte, et al., The biochemical pathway of neurofibrillary degeneration in aging and 726
645 Alzheimer's disease. *Neurology* **52**, 1158–1165 (1999). 727
- 646 7. H Braak, E Braak, Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 728
647 **82**, 239–259 (1991). 729
- 648 8. RJ Bateman, et al., Clinical and biomarker changes in dominantly inherited alzheimer's disease. 730
649 *New Engl. J. Medicine* **367**, 795–804 (2012) PMID: 22784036. 731
- 650 9. J Cummings, G Lee, A Ritter, M Sabbagh, K Zhong, Alzheimer's disease drug development 732
651 pipeline: 2019. *Alzheimers Dement (N Y)* **5**, 272–293 (2019). 733
- 652 10. R Howard, KY Liu, Questions EMERGE as Biogen claims aducanumab turnaround. *Nat Rev* 734
653 *Neurol* **16**, 63–64 (2020). 735
- 654 11. LS Honig, et al., Trial of Solanezumab for Mild Dementia Due to Alzheimer's Disease. *N.* 736
655 *Engl. J. Med.* **378**, 321–330 (2018). 737
- 656 12. MF Egan, et al., Randomized Trial of Verubecestat for Prodromal Alzheimer's Disease. *N.* 738
657 *Engl. J. Med.* **380**, 1408–1420 (2019). 739
- 658 13. AM Wessels, et al., Efficacy and Safety of Lanabecestat for Treatment of Early and Mild 740
659 Alzheimer Disease: The AMARANTH and DAYBREAK-ALZ Randomized Clinical Trials. 741
660 *JAMA Neurol* **77**, 199–209 (2020). 742
- 661 14. D Henley, et al., Preliminary Results of a Trial of Atabecestat in Preclinical Alzheimer's Disease. 743
662 *N. Engl. J. Med.* **380**, 1483–1485 (2019). 744
- 663 15. PS Aisen, et al., What Have We Learned from Expedition III and EPOCH Trials? Perspective 745
664 of the CTAD Task Force. *J Prev Alzheimers Dis* **5**, 171–174 (2018). 746
- 665 16. G Klein, et al., Gantenerumab reduces amyloid- β plaques in patients with prodromal to moderate 747
666 alzheimer's disease: a pet substudy interim analysis. *Alzheimer's Res. & Ther.* **11**, 101 748
667 (2019). 749
- 668 17. HM Fonteijn, et al., An event-based model for disease progression and its application in familial 750
669 Alzheimer's disease and Huntington's disease. *NeuroImage* **60**, 1880–1889 (2012). 751
- 670 18. BM Jedynak, et al., A computational neurodegenerative disease progression score: method and 752
671 results with the Alzheimer's disease Neuroimaging Initiative cohort. *NeuroImage* **63**, 753
672 1478–1486 (2012). 754
- 673 19. NP Oxtoby, et al., Data-Driven Sequence of Changes to Anatomical Brain Connectivity in Sporadic 755
674 Alzheimer's Disease. *Front Neurol* **8**, 580 (2017). 756
- 675 20. J Schiratti, S Allassonnière, O Colliot, S Durleman, Learning spatiotemporal trajectories from 757
676 manifold-valued longitudinal data in NIPS. pp. 2404–2412 (2015). 758
- 677 21. CA Nader, N Ayache, P Robert, M Lorenzi, ADN Initiative, Monotonic gaussian process for 759
678 spatio-temporal disease progression modeling in brain imaging data. *NeuroImage* **205** 760
679 (2020). 761
- 680 22. M Bilgel, B Jedynak, DF Wong, SM Resnick, JL Prince, Temporal Trajectory and Progression 762
681 Score Estimation from Voxelwise Longitudinal Imaging Measures: Application to Amyloid 763
682 Imaging. *Inf Process. Med Imaging* **24**, 424–436 (2015). 764
- 683 23. MC Donohue, et al., Estimating long-term multivariate progression from short-term data. 765
684 *Alzheimer's & Dementia* **10**, S400 – S410 (2014). 766
- 685 24. Y Iturria-Medina, et al., Early role of vascular dysregulation on late-onset Alzheimer's disease 767
686 based on multifactorial data-driven analysis. *Nat Commun* **7**, 11934 (2016). 768
- 687 25. M Lorenzi, M Filippone, GB Frisoni, DC Alexander, S Ourselin, Probabilistic disease progres- 769
688 sion modeling to characterize diagnostic uncertainty: Application to staging and prediction in 770
689 Alzheimer's disease. *NeuroImage* **190**, 56–68 (2019). 771
- 690 26. RV Marinescu, et al., DIVE: A spatiotemporal progression model of brain pathology in neuro- 772
691 degenerative disorders. *NeuroImage* **192**, 166–177 (2019). 773
- 692 27. AL Young, et al., A data-driven model of biomarker changes in sporadic Alzheimer's disease. 774
693 *Brain* **137**, 2564–2577 (2014). 775
- 694 28. NP Oxtoby, et al., Data-driven models of dominantly-inherited Alzheimer's disease progres- 776
695 sion. *Brain* **141**, 1529–1544 (2018). 777
- 696 29. D Li, S Iddi, WK Thompson, MC Donohue, Bayesian latent time joint mixed effect models for 778
697 multicohort longitudinal data. *Stat Methods Med Res* **28**, 835–845 (2019). 779
- 698 30. W Hao, A Friedman, Mathematical model on Alzheimer's disease. *BMC Syst Biol* **10**, 108 780
699 (2016). 781
31. JR Petrella, W Hao, A Rao, PM Doraiswamy, Computational Causal Modeling of the Dynamic 782
700 Biomarker Cascade in Alzheimer's Disease. *Comput. Math Methods Med* **2019**, 6216530 701
702 (2019). 703
32. Y Iturria-Medina, FM Carbonell, RC Sotero, F Chouinard-Decorte, AC Evans, Multifactorial 704
705 causal model of brain (dis)organization and therapeutic intervention: Application to 706
707 alzheimer's disease. *NeuroImage* **152**, 60 – 77 (2017). 708
33. S Garbarino, M Lorenzi, Modeling and Inference of Spatio-Temporal Protein Dynamics Across 709
710 Brain Networks in *IPMI 2019 - 26th International Conference on Information Processing in* 711
712 *Medical Imaging*, LNCS. (Springer, Hong-Kong, China), Vol. 11492, pp. 57–69 (2019). 713
34. L Antelmi, N Ayache, P Robert, M Lorenzi, Sparse Multi-Channel Variational Autoencoder for 714
715 the Joint Analysis of Heterogeneous Data in *ICML 2019 - 36th International Conference on* 716
717 *Machine Learning*. (Long Beach, United States), (2019). 718
35. LM Shaw, et al., Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging 719
720 initiative subjects. *Ann. Neurol.* **65**, 403–413 (2009). 721
36. DM Cash, et al., Assessing atrophy measurement techniques in dementia: Results from the 722
723 MIRIAD atrophy challenge. *NeuroImage* **123**, 149–164 (2015). 724
37. N Schuff, et al., MRI of hippocampal volume loss in early Alzheimer's disease in relation to 725
726 ApoE genotype and biomarkers. *Brain* **132**, 1067–1077 (2009). 727
38. RV Marinescu, DC Alexander, P Golland, Brainpainter: A software for the visualisation of 728
729 brain structures, biomarkers and associated pathological processes. *CoRR abs/1905.08627* 730
731 (2019). 732
39. D Gamberger, N Lavrač, S Srivatsa, RE Tanzi, PM Doraiswamy, Identification of clusters of 733
734 rapid and slow decliners among subjects at risk for Alzheimer's disease. *Sci Rep* **7**, 6763 735
736 (2017). 737
40. R Kochhann, JS Varela, CSM Lisboa, MLF Chaves, The Mini Mental State Examination: 738
739 Review of cutoff points adjusted for schooling in a large Southern Brazilian sample. *Dement* 740
741 *Neuropsychol* **4**, 35–41 (2010). 742
41. RA Sperling, CR Jack, PS Aisen, Testing the right target and right drug at the right stage. *Sci* 743
744 *Transl Med* **3**, 111cm33 (2011). 745
42. CC Rowe, et al., Amyloid imaging results from the Australian Imaging, Biomarkers and 746
747 Lifestyle (AIBL) study of aging. *Neurobiol. Aging* **31**, 1275–1283 (2010). 748
43. H Zetterberg, SC Burnham, Blood-based molecular biomarkers for alzheimer's disease. *Mol.* 749
750 *brain* **12**, 26–26 (2019). 751
44. K Blennow, H Hampel, M Weiner, H Zetterberg, Cerebrospinal fluid and plasma biomarkers 752
753 in Alzheimer disease. *Nat Rev Neurol* **6**, 131–144 (2010). 754
45. S Westwood, et al., Blood-Based Biomarker Candidates of Cerebral Amyloid Using PiB PET in 755
756 Non-Demented Elderly. *J. Alzheimers Dis.* **52**, 561–572 (2016). 757
46. S Gauthier, et al., Combination Therapy for Alzheimer's Disease: Perspectives of the EU/US 758
759 CTAD Task Force. *J Prev Alzheimers Dis* **6**, 164–168 (2019). 760
47. SK Kaufman, K Del Tredici, TL Thomas, H Braak, MI Diamond, Tau seeding activity begins in 761
762 the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in alzheimer's 763
764 disease and part. *Acta Neuropathol.* **136**, 57–67 (2018). 764
48. MJ Pontecorvo, et al., A multicentre longitudinal study of flortaucipir (18F) in normal ageing, 765
766 mild cognitive impairment and Alzheimer's disease dementia. *Brain* **142**, 1723–1735 (2019). 766
49. J Cummings, et al., Anti-Tau Trials for Alzheimer's Disease: A Report from the EU/US/CTAD 767
768 Task Force. *J Prev Alzheimers Dis* **6**, 157–163 (2019). 768
50. M Safieh, AD Koczy, DM Michaelson, ApoE4: an emerging therapeutic target for 769
770 Alzheimer's disease. *BMC Medicine* **17** (2019). 770
51. M Lorenzi, et al., Susceptibility of brain atrophy to trib3 in alzheimer's disease, evidence from 771
772 functional prioritization in imaging genetics. *Proc. Natl. Acad. Sci.* **115**, 3162–3167 (2018). 772
52. J Kim, JM Basak, DM Holtzman, The role of apolipoprotein E in Alzheimer's disease. *Neuron* 773
774 **63**, 287–303 (2009). 774
53. R Sivera, et al., Voxel-based assessments of treatment effects on longitudinal brain changes 775
776 in the multidomain alzheimer preventive trial cohort. *Neurobiol. Aging* **94**, 50 – 59 (2020). 775
54. M Lorenzi, X Pennec, GB Frisoni, N Ayache, Disentangling normal aging from alzheimer's 776
777 disease in structural magnetic resonance images. *Neurobiol. Aging* **36**, S42 – S52 (2015) 776
777 Novel Imaging Biomarkers for Alzheimer's Disease and Related Disorders (NIBAD). 777
55. M Reuter, NJ Schmansky, HD Rosas, B Fischl, Within-subject template estimation for unbi- 778
779 ased longitudinal image analysis. *NeuroImage* **61**, 1402–1418 (2012). 778
56. RS Desikan, et al., An automated labeling system for subdividing the human cerebral cortex 779
780 on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006). 780
57. Z Ghahramani, M Beal, Graphical models and variational methods in *Advanced mean field* 781
782 *methods: theory and practice*, Neural Information Processing, eds. M Opper, D Saad. (MIT), 781
783 (2001). 782
58. DP Kingma, M Welling, Auto-encoding variational bayes. *CoRR abs/1312.6114* (2013). 783
784 784
59. DP Kingma, J Ba, Adam: A method for stochastic optimization (2014) cite 785
786 arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Con- 785
787 ference for Learning Representations, San Diego, 2015. 786
60. A Paszke, et al., Pytorch: An imperative style, high-performance deep learning library in 787
788 *Advances in Neural Information Processing Systems* **32**, eds. H Wallach, et al. (Curran 787
789 Associates, Inc.), pp. 8024–8035 (2019). 788
61. TQ Chen, Y Rubanova, J Bettencourt, DK Duvenaud, Neural ordinary differential equations in 789
790 *Advances in Neural Information Processing Systems* **31**, eds. S Bengio, et al. (Curran 789
791 Associates, Inc.), pp. 6571–6583 (2018). 790
62. H Akaike, *Information Theory and an Extension of the Maximum Likelihood Principle.* 791
792 (Springer New York, New York, NY), pp. 199–213 (1998). 792

1

2 **Supplementary Information for**

3 **Simulating the outcome of amyloid treatments in Alzheimer's Disease from multi-modal** 4 **imaging and clinical data**

5 **Clément Abi Nader, Nicholas Ayache, Giovanni B. Frisoni, Philippe Robert and Marco Lorenzi**

6 **Corresponding Author: Clément Abi Nader.**

7 **E-mail: clement.bi-nader@inria.fr**

8 **This PDF file includes:**

- 9 Figs. S1 to S5
- 10 Tables S1 to S3
- 11 SI References

12 **Time-shift comparison and validation.** We compared our estimated disease severity (Figure 2 Panel II in the manuscript) with
 13 the one obtained applying the monotonic Gaussian Process (GP) model of (1) from the state-of-the-art (Figure S1A). While
 14 both methods estimate significant time differences when going from healthy to pathological stages, our approach captures a
 15 larger temporal variability for both earlier and later stages of the disease, as shown in Figure S1B, highlighting a stronger
 16 separability across clinical stages.

17
 18 We also assessed the model on an independent testing cohort from the ADNI composed of 130 NL stable, 10 NL converters, 125
 19 MCI stable, 7 MCI converters, and 12 AD subjects which were not necessarily amyloid positive. It is important to note that no
 20 PET-FDG data was available for these subjects. We provide in Table S1 socio-demographic and clinical information for the
 21 testing cohort across the different clinical groups. Despite the fact that no FDG data was used to estimate the disease severity,
 22 we observe in Figure S2 that the method still exhibits good separating performances between clinical stages, coherently with
 23 the clinical status of the testing individuals.

Estimated disease severity across clinical stages
(monotonic GP)

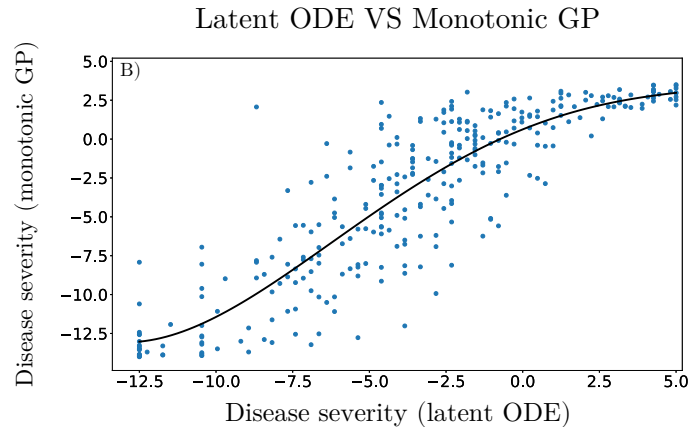
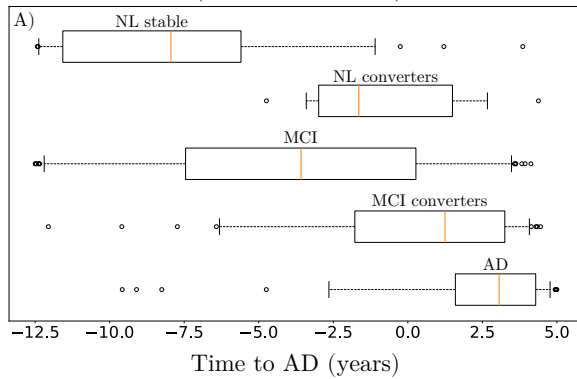


Fig. S1. A: Distribution of the disease severity estimated by the monotonic GP method (1) on the training set. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer’s dementia. B: Comparison of the disease severity estimated by our method (denoted as latent ODE) with respect to the one estimated by the monotonic GP.

Table S1. 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)

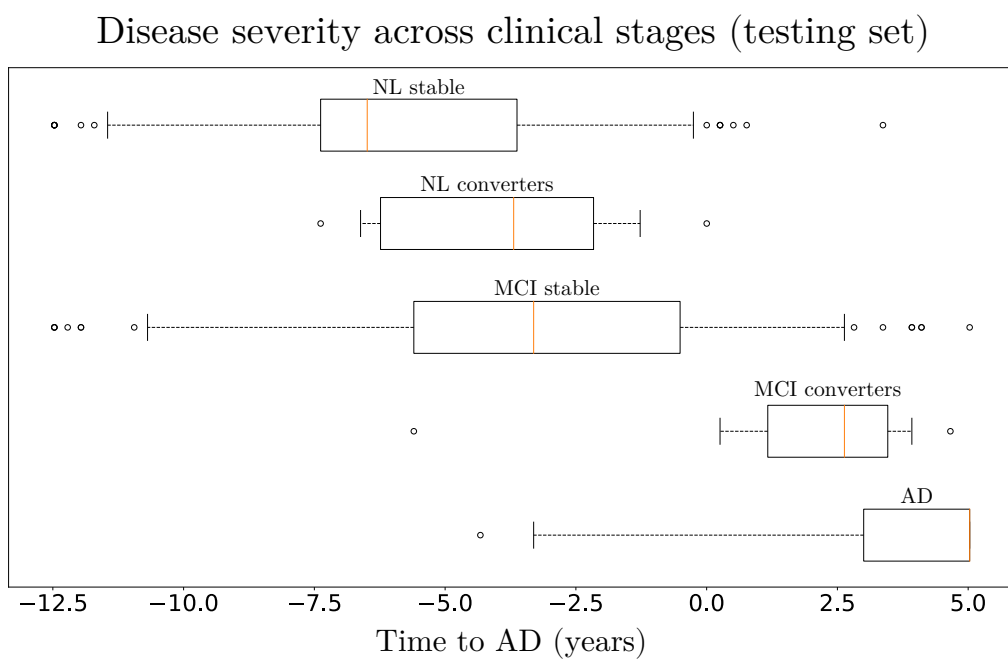


Fig. S2. Distribution of the disease severity estimated for the subjects of the testing set, relatively to the long-term dynamics of Figure 2 Panel II) left in the manuscript. NL: normal individuals, MCI: mild cognitive impairment, AD: Alzheimer's dementia.

24 **Clinical scores progression.** We provide in Figure S3 the estimated progression of the RAVLT immediate and RAVLT forgetting
 25 scores based on the z-score trajectories of Figure 2 Panel II) (left) in the manuscript.

Clinical outcomes evolution

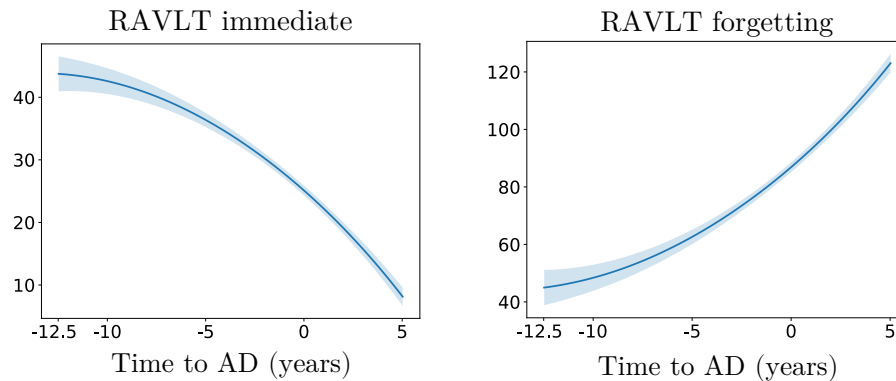


Fig. S3. Modelled long-term evolution of RAVLT immediate and RAVLT forgetting scores. Shaded areas represent the standard deviation of the average trajectory.

26 **Effect of intervention for RAVLT immediate and RAVLT forgetting.** Similarly to Figure 4 in the manuscript, Figure S4 shows the
 27 change in statistical power depending on intervention time and sample size for the RAVLT immediate and RAVLT forgetting.
 28 We also provide in Table S2 the estimated improvement for RAVLT immediate and RAVLT forgetting at conversion with a
 29 sample size of 100 subjects per arm for both 100% and 50% amyloid lowering depending on the intervention time.

Amyloid lowering intervention 100% Amyloid lowering intervention 50%

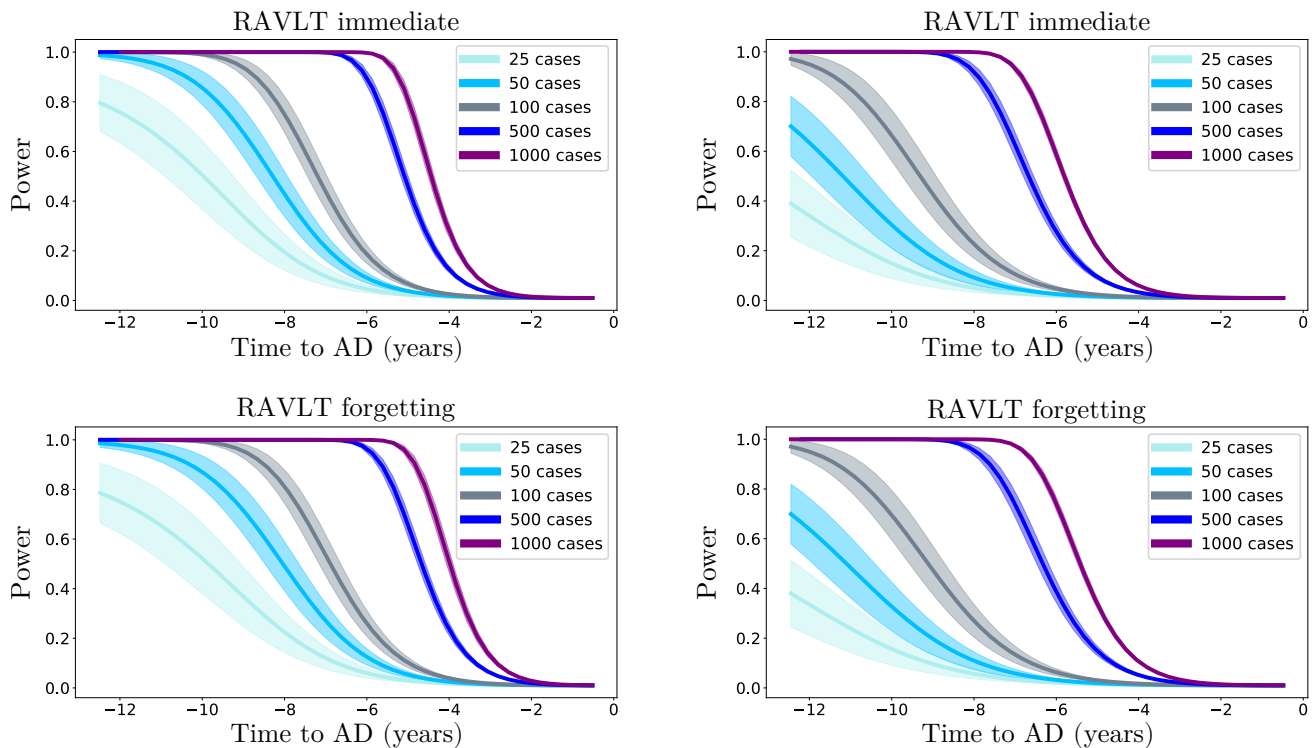


Fig. S4. Modelled long-term evolution of RAVLT immediate and RAVLT forgetting scores. Shaded areas represent the standard deviation of the average trajectory.

Table S2. Estimated mean (standard deviation) improvement of RAVLT immediate and RAVLT forgetting 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, 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
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
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)

30 Simulated clinical endpoints. We provide in Table S3 the estimated values for each clinical score at predicted conversion time
31 for the normal progression case when performing the simulations presented in Section Simulating clinical intervention.

Table S3. 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								
Years to AD \ Score	-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								
Years to AD \ Score	-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)

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

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

33 Given that:

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

34 We obtain:

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

35 **KL divergence.** We have that:

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

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

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

37 **Graphical model.** Figure S5 below provides the graphical model illustrating the method presented in Materials and Methods.

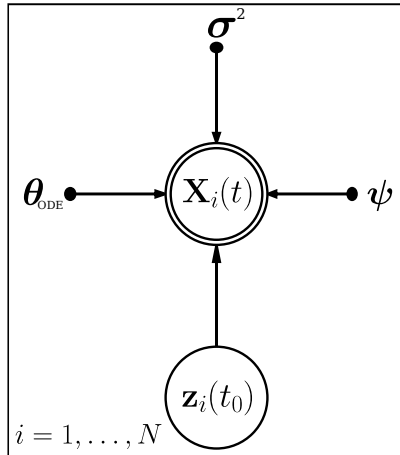


Fig. S5. Graphical model of the proposed method.

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

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}
-

39 References

- 40 1. M Lorenzi, M Filippone, GB Frisoni, DC Alexander, S Ourselin, Probabilistic disease progression modeling to characterize
 41 diagnostic uncertainty: Application to staging and prediction in Alzheimer’s disease. *NeuroImage* **190**, 56–68 (2019).