



**HAL**  
open science

## **SimulAD: A dynamical model for personalized simulation and disease staging in Alzheimer's disease**

Clément Abi Nader, Federica Ribaldi, Giovanni B Frisoni, Valentina Garibotto, Philippe Robert, Nicholas Ayache, Marco Lorenzi

► **To cite this version:**

Clément Abi Nader, Federica Ribaldi, Giovanni B Frisoni, Valentina Garibotto, Philippe Robert, et al..  
SimulAD: A dynamical model for personalized simulation and disease staging in Alzheimer's disease.  
Neurobiology of Aging, 2022, 113, pp.73-83. 10.1016/j.neurobiolaging.2021.12.015 . hal-03514292

**HAL Id: hal-03514292**

**<https://hal.inria.fr/hal-03514292>**

Submitted on 6 Jan 2022

**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.

# SimulAD: A dynamical model for personalized simulation and disease staging in Alzheimer's disease

Clément Abi Nader <sup>\*1</sup>, Federica Ribaldi<sup>2</sup>, Giovanni B. Frisoni<sup>2</sup>, Valentina Garibotto<sup>3,4</sup>, Philippe Robert<sup>5</sup>, Nicholas Ayache<sup>1</sup> and Marco Lorenzi<sup>1</sup>

<sup>1</sup>Université Côte d'Azur, Inria Sophia Antipolis, Epione Research Project, France.

<sup>2</sup>Memory Clinic and LANVIE-Laboratory of Neuroimaging of Aging, Hospitals and University of Geneva, Geneva, Switzerland.

<sup>3</sup>Faculty of Medicine, Geneva University, Geneva, Switzerland.

<sup>4</sup>Nuclear Medicine, Geneva University, Geneva, Switzerland.

<sup>5</sup>Université Côte d'Azur, CoBTeK lab, MNC3 program, France.

Tuesday 16<sup>th</sup> November, 2021

## Abstract

*SimulAD is a computational disease progression model (DPM) originally developed on the ADNI database to simulate the evolution of clinical and imaging markers characteristic of AD, and to quantify the disease severity (DS) of a subject. In this work, we assessed the validity of this estimated DS, as well as the generalization of the DPM, by applying SimulAD on a new cohort from the Geneva Memory Center (GMC). The differences between the estimated DS of healthy, mild cognitive impairment and AD dementia groups were statistically significant ( $p$ -values  $< 0.05$ ;  $d \geq 0.8$ ). DS correlated with MMSE ( $\rho = -0.55$ ), hippocampal atrophy ( $\rho = -0.62$ ), glucose hypometabolism ( $\rho = -0.67$ ), amyloid burden ( $\rho = 0.31$ ) and tau deposition ( $\rho = 0.62$ ) ( $p$ -values  $< 0.01$ ). Based on the dynamics estimated on the ADNI cohort, we simulated a DPM for the subjects of the GMC cohort. The difference between the temporal evolution of similar biomarkers simulated on the ADNI and GMC cohorts remained below 10%. This study illustrates the robustness and good generalization of SimulAD, highlighting its potential for clinical and pharmaceutical studies.*

**Keywords**— Alzheimer's disease, Disease progression models, Clinical trials, Biomarkers

---

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

---

# 1 Introduction

Alzheimer’s disease (AD) is a neurodegenerative disorder whose evolution has been hypothesized to follow a cascade of events (Jack, Knopman, et al. 2013). Deposition of the beta-amyloid protein in the brain cortex is believed to initiate this cascade, and to subsequently cause the aggregation of hyperphosphorylated tau protein in neurofibrillary tangles. This is followed by a process of neurodegeneration (i.e glucose hypometabolism and gray matter atrophy) ultimately leading to dementia. An inherent difficulty in diagnosing AD is that patients go through a long asymptomatic phase spanning approximately 10 to 20 years (Sperling et al. 2011) before showing clinical symptoms. To provide a biological assessment of the disease, AD has been recently defined as a pathology characterized by three main biomarkers categories, namely: amyloid, tau and neurodegeneration (Jack, Bennett, et al. 2018). These three biomarkers can be measured thanks to imaging techniques, such as Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET), or in the case of amyloid and tau, also by lumbar puncture and blood collection. Monitoring these biomarkers is paramount in order to track the disease progression, and to potentially facilitate prevention or assessment of drug efficacy.

In the past years, the proliferation of studies collecting large amounts of biomarkers, combined with the growth of machine learning, fostered the development of computational models for automated AD diagnosis. For instance, many studies focused on the development of data-driven approaches for automatic assessment of clinical diagnosis (Davatzikos et al. 2009; Arbabshirani et al. 2017; Falahati et al. 2014). Based on the sole analysis of imaging-derived data, these methods showed that it is possible to automatically identify healthy controls, subjects with mild cognitive impairment, and patients suffering from AD dementia, some of them reporting results comparable to diagnosis rates obtained by expert physicians (Klöppel et al. 2008). However, most of these approaches have

---

39 been exclusively developed to solve a predictive task, and generally don't allow to understand the  
40 mechanisms relating the different biomarkers throughout AD evolution. As these mechanisms still  
41 remain partially unknown, different methods known as disease progression models were therefore  
42 introduced in order to estimate, in a data-driven fashion, the long-term progression of biomarkers  
43 (Jedynak et al. 2012). Due to the lack of an absolute time-line describing AD evolution, these  
44 models usually assume that the disease is characterized by monotonic changes, such that the  
45 modelled biomarkers steadily evolve from normal to pathological values. This assumption allows  
46 to reconstruct a time-line on which we can track the disease progression (Lorenzi et al. 2017).  
47 Moreover, these methods can be applied to a variety of data types, such as cortical and subcortical  
48 shapes (Marinescu et al. 2019) or volumetric images (Khanal, Lorenzi, et al. 2016; Khanal, Ayache,  
49 et al. 2017), thus offering a fine-grained spatial description of the changes affecting the brain. These  
50 models also allow to automatically assess the individual disease severity by comparing the clinical  
51 and imaging measurements of a given subject to the estimated disease progression. This latter  
52 capability of disease progression models is usually referred as disease staging. Ultimately, these  
53 approaches could potentially be used for identifying individuals at risk of cognitive decline, or for  
54 assessing drug efficacy in clinical trials.

55

56 Since these statistical models have been mostly developed on publicly available research datasets,  
57 such as the one provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI), their gener-  
58 alization to external cohorts from memory clinics still requires additional testing and validation  
59 (Mendelson et al. 2017). As clinical cohorts may fundamentally differ from the ADNI one, whether  
60 it be in terms of data acquisition or study population, automated diagnosis pipelines usually show  
61 a prominent decrease in performances. Regarding disease progression models, it is conceivable  
62 that biomarkers' trajectories estimated solely via the analysis of a single cohort may not be fully

---

63 representative of the natural disease course. This aspect would question the generalization of  
64 the progression model when tested on subjects from independent clinical cohorts. It is there-  
65 fore essential to assess the generalization of this kind of models on independent datasets, with  
66 respect to their specific biases which can encompass a broad range of differences between cohorts  
67 such as data acquisition, missing data or data heterogeneity. These differences need to be ad-  
68 dressed to finally deploy disease progression models in a practical clinical setting (Castro et al. 2020).

69  
70 Recently, the event-based-model (EBM) (Fontejn et al. 2012) and the Discriminative EBM  
71 (DEBM) (Venkatraghavan et al. 2019) underwent an extensive evaluation effort. These approaches  
72 model AD progression as a sequence of events representing the transition of a set of biomarkers  
73 from a normal to an abnormal state. Both EBM and DEBM have been applied on subjects from  
74 independent cohorts, providing accurate patients staging (Archetti et al. 2019). However, the kind of  
75 progression model estimated by these methods presents important limitations. First, both approaches  
76 are based on the simplistic assumption describing the pathological progression as a discrete sequence  
77 of biomarkers transitions from normal to abnormal states, which doesn't reflect the continuous  
78 nature of the changes affecting the brain during the disease. Second, they allow the analysis of  
79 summary measures only, such as regional brain uptake of grey matter density values, and thus don't  
80 enable the fine-grained quantification of the spatial patterns of changes associated with the disease.  
81 Third, while these two models inform us about the sequence of events characterizing AD, they don't  
82 provide insights about the dynamical interplay between biomarkers. Investigating such interactions  
83 would allow a deeper understanding of how the pathological processes at stake during the disease  
84 affect each other. Fourth, the EBM and the DEBM don't allow to simulate hypothetical scenarios of  
85 disease progression. Yet, such capability could be used to assess the effect of drug intervention on  
86 the disease evolution *in silico*, which could help planning and monitoring clinical trials.

---

87

88 To address these limitations, more refined approaches to disease progression modeling such  
89 as the Subtype and Stage Inference (SuStaIn) (Young et al. 2018) method have been proposed.  
90 This method reformulates the EBM and DEBM in order to model the disease progression over a  
91 continuous timeline, thus tackling the main limitation of the EBM and the DEBM which described  
92 the pathological progression as a sequence of discrete events. In addition, we note that SuStaIn  
93 allows to identify subtypes of disease progression, thus addressing the problem of AD heterogeneity  
94 which is one of the major challenges for the development of personalized disease progression models  
95 and their application in clinical practice for staging patients. Besides SuStaIn, many others disease  
96 progression models have been proposed to enable the fine-grained description of the pathological  
97 evolution in space and time (Marinescu et al. 2019; Koval et al. 2018; Bilgel et al. 2016). Within  
98 this context, SimulAD (Abi Nader et al. 2021) is a recent method allowing the analysis of clinical  
99 scores and multivariate imaging data extracted from MRI and PET scans to estimate a continuous  
100 spatio-temporal model of disease progression. Compared to the EBM, DEBM and SuStaIn, this  
101 approach offers a higher resolution for the imaging biomarkers, allowing to track the evolution of  
102 regional changes affecting the brain along a continuous temporal scale describing the disease course  
103 over 30 years. The trajectories estimated for clinical and imaging markers can subsequently be  
104 used as a reference to assess the individual disease severity, by locating subjects along the temporal  
105 scale describing the disease evolution. Moreover, SimulAD estimates the dynamical relationships  
106 between key biomarkers at stake during AD progression, namely: amyloid deposition, glucose  
107 hypometabolism, cerebral atrophy, cognitive and behavioural decline. Based on these relationships,  
108 the model can be applied to simulate the personalized evolution of any patient or group of patients  
109 only from the knowledge of their baseline clinical and imaging measurements. Finally, SimulAD  
110 enables to assess the impact of therapeutic intervention, such as anti-amyloid treatment, on cognitive

---

111 outcomes depending on the intervention time (Abi Nader et al. 2021).

112

113 While SimulAD was previously trained on the ADNI cohort, its generalization to independent  
114 datasets has not been evaluated. As it generally holds for statistical and machine learning models, one  
115 of the main difficulty for the generalization of SimulAD to independent cohorts lies in the challenge  
116 of accounting for potential missing data, heterogeneity due to different acquisition protocols or even  
117 data incompatibility between cohorts. Yet, evaluating the generalization of SimulAD to independent  
118 cohorts is essential to demonstrate the reliability of the dynamics allowing to personalize the models  
119 of disease progression, as well as the validity of the approach for providing accurate disease staging.

120

121 Within this context, we assess in this study the generalization capabilities of SimulAD. To this end,  
122 we test the robustness and reliability of this approach when applied to an independent dataset from  
123 a memory clinic, namely the Geneva Memory Center (GMC). This cohort includes patients with  
124 cognitive complaints, who underwent a baseline clinical and neuropsychological evaluation, MRI,  
125 amyloid-PET, (18)F-fluorodeoxyglucose-PET (FDG-) and (18)F-flortaucipir-PET (tau-) scans. Our  
126 evaluation procedure relies on three key aspects: (i) Development of a pre-processing pipeline  
127 allowing to apply SimulAD on the GMC cohort; (ii) Assessment of SimulAD validity for individual  
128 disease staging. (iii) Evaluation of the reliability of the progression of imaging and clinical markers  
129 estimated by SimulAD.

## 130 **2 Material and methods**

131 In this work subjects were divided in five clinical groups: cognitively healthy (NL stable), individuals  
132 diagnosed with mild cognitive impairment (MCI stable), patient suffering from Alzheimer’s disease

---

133 dementia (AD dementia), subjects progressing from NL to MCI or AD dementia (NL converters),  
134 and finally subjects progressing from MCI to AD dementia (MCI converters).

135

## 136 **2.1 Experimental cohorts**

137 The ADNI and GMC cohorts respectively included 442 and 93 subjects. In Table 1, we provide  
138 socio-demographic information across clinical groups for these datasets. The clinical spectrum of  
139 the cohorts spans a broad range of cognitive severity, from healthy to moderate and severe dementia.  
140 Conversion to AD dementia was determined using the last available follow-up information. In both  
141 the ADNI and GMC cohorts, MCI converters are subjects who were diagnosed as MCI at baseline  
142 and subsequently progressed to AD. In the case of the ADNI database, diagnosis was established  
143 using the DX column from the ADNIMERGE file (<https://adni.bitbucket.io/index.html>), which  
144 reflects the standard ADNI clinical assessment based on Wechsler Memory Scale, Mini-Mental  
145 State Examination, and Clinical Dementia Rating. Concerning the subjects from the GMC cohort,  
146 diagnosis was established after that the patients underwent neuro-psychological tests under the  
147 supervision of a psychologist. MCI converters from the ADNI database had a mean follow-up of 9  
148 years with a standard deviation of 1.5 years, while the ones from the GMC cohort were followed in  
149 average 2.4 years with a standard deviation of 1.7. All the participants were amyloid positive at  
150 baseline. In the case of the ADNI cohort, “amyloid positive” subjects are patients whose amyloid  
151 level in the CSF was below the nominal cutoff of 192 pg/ml (Gamberger et al. 2017) either at baseline  
152 or during any follow-up visit. Concerning the GMC cohort, “amyloid positivity” was evaluated  
153 using visual assessment performed by an expert nuclear medicine physician and following the tracer  
154 manufacturers guidelines. Concerning the ADNI cohort, multi-modal biomarkers consisting of  
155 neuropsychological tests and measures derived from MRI, FDG-PET, and Amyloid-PET scans



---

156 were collected at baseline and during follow-up visits. In the case of the GMC cohort, each  
157 participant underwent the Mini-Mental State Examination (MMSE), as well as an MRI, FDG-PET  
158 and amyloid-PET scan at baseline. In addition, a tau-PET scan was acquired for 50 subjects  
159 from this cohort at baseline. During subsequent visits, only the MMSE was assessed. Summary  
160 statistics about clinical and imaging-derived information across clinical groups are reported in Table 1.

161

## 162 **2.2 Image preprocessing**

163 We derived volumes of gray matter density in a standard anatomical space by relying on the  
164 longitudinal pipeline of Freesurfer (Reuter et al. 2012). Regional gray matter density was extracted  
165 from the Desikan-Killiany parcellation (Desikan et al. 2006). amyloid-PET, FDG-PET and tau-PET  
166 images were registered to their corresponding T1-MRI acquisition, and normalized to the cerebellum  
167 uptake. Regional amyloid load, glucose metabolism and tau burden were computed thanks to the  
168 PetSurfer software (Greve et al. 2014). For every imaging modality we discarded white-matter,  
169 ventricular, and cerebellar regions, thus obtaining 82 regions that were averaged across hemispheres.

## 170 **2.3 Modeling framework**

171 SimulAD is based on the hypothesis that AD evolution can be mathematically modelled by a set of  
172 key biomarkers following a dynamical system. These biomarkers are namely *clinical scores*, *gray*  
173 *matter atrophy*, *amyloid load* and *glucose metabolism*. This assumption has two consequences:  
174 the first one is that, at any given time, AD severity is uniquely associated with the values of these  
175 biomarkers. The second one is that past and futures states of the disease can be computed from the  
176 current ones thanks to mathematical relationships linking the biomarkers evolutions.

177

Table 1: Baseline characteristics of the ADNI and GMC cohorts. Average values, standard deviation in parenthesis. Acronyms: ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer’s dementia; FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging; SUVR: Standardized Uptake Value Ratio; MMSE: Mini Mental State Examination; Tau: (18)F-flortaucipir PET imaging. Converters are NL patients progressing to MCI or AD dementia, or MCI individuals progressing to AD dementia. Hippocampal volume: extracted with Freesurfer. Amyloid burden: voxel-number weighted average of the amyloid uptake in the frontal, anterior/posterior cingulate, lateral parietal, and lateral temporal regions normalized to the cerebellum. Early amyloid: voxel-number weighted average of the uptake extracted from the early-phase (6 min) of amyloid-PET in the frontal, anterior/posterior cingulate, lateral parietal, and lateral temporal regions normalized to the cerebellum. Glucose metabolism: voxel-number weighted average of the FDG uptake in the angular, temporal, and posterior cingulate cortex normalized to the cerebellum. Tau burden: voxel-number weighted average of the tau uptake in the entorhinal, amygdala, parahippocampal, fusiform, inferior temporal, and middle temporal regions normalized to the cerebellum. Amyloid-corrected and FDG-corrected indicate the values obtained after performing the data adjustment presented in Section 2.6. Missing data for 36<sup>a</sup> and 43<sup>b</sup> subjects.

Group	NL stable		MCI stable		MCI converters		AD dementia	
	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI
Cohort								
N	23	71	28	131	25	105	17	102
Female (%)	61	62	61	37	65	47	53	45
Age (years)	69.1 (7.5)	73.7( 6.0)	74.3 (6.5)	72.0 (7.6)	73.8 (4.6)	72.6 (6.7)	70.4 (11.1)	73.6 (8.2)
Education (years)	17.3 (3.9)	16.3 (2.5)	14.0 (3.1)	16.3 (2.7)	12.5 (4.7)	16.2 (2.6)	11.6 (3.8)	15.6 (2.5)
MMSE	28.5 (1.0)	29.2 (1.1)	25.2 (3.1)	28.1 (1.8)	23.7 (4.3)	26.4 (2.7)	18.2 (6.5)	23.0 (2.1)
Hippocampus (mm <sup>3</sup> )	4271 (435)	4047 (455)	3621 (534)	3907 (541)	3634 (448)	3517 (515)	3409 (436)	3401 (472)
Amyloid (SUVR)	0.67 (0.13)	0.74 (0.17)	0.96 (0.17)	0.79 (0.18)	0.93 (0.14)	0.92 (0.17)	0.89 (0.12)	0.92 (0.16)
Amyloid-corrected (SUVR)	0.70 (0.12)	/	0.93 (0.13)	/	0.91 (0.13)	/	0.89 (0.11)	/
Early amyloid (SUVR)	0.53 (0.03)	/	0.50 (0.03)	/	0.47 (0.02)	/	0.45 (0.03)	/
FDG <sup>a</sup> (SUVR)	0.61 (0.05)	0.62 (0.04)	0.52 (0.04)	0.61 (0.06)	0.52 (0.04)	0.56 (0.06)	0.49 (0.06)	0.52 (0.05)
FDG-corrected SUVR	0.57 (0.03)	/	0.54 (0.04)	/	0.53 (0.04)	/	0.50 (0.05)	/
Tau <sup>b</sup> (SUVR)	0.60 (0.06)	/	0.81 (0.20)	/	0.94 (0.27)	/	1.21 (0.34)	/

---

178 To estimate the complex relationships between high-dimensional imaging and clinical measures, the  
179 model first transforms baseline neuropsychological assessments and measures derived from MRI,  
180 amyloid-PET and FDG-PET data in a set of four corresponding z-scores. The transformation consists  
181 in a weighted average of the measurements derived from each type of data modality (i.e regional  
182 grey matter measurements in the case of *atrophy*). The obtained z-scores are respectively denoted  
183  $z^{cli}$ ,  $z^{atr}$ ,  $z^{amy}$   $z^{met}$ , and describe the overall pathological status of an individual. We hypothesize  
184 that these four z-scores are related by a set of relationships driving the disease progression, which  
185 are mathematically modelled by a system of Ordinary Differential Equations (ODEs). This system  
186 of ODEs provides us with an interaction rule that describes how the z-scores jointly evolve over time.  
187 The parameters controlling the system of ODEs are optimized such that the predicted evolution  
188 of the z-scores best matches the available follow-up clinical and imaging measurements of each  
189 individual. An overview of the model is provided in Supplementary Section 1. Further details  
190 about model optimization and mathematical transformations allowing to map the z-scores and raw  
191 measures are given in (Abi Nader et al. 2021).

192

193 **Trajectory modelling.** Thanks to this mathematical formulation, SimulAD can be used to simulate  
194 the progression of changes characterizing AD by considering the subjects diagnosed with AD  
195 dementia, and for whom we compute corresponding z-scores based on their baseline measures.  
196 Relying on the estimated set of relationships between z-scores we follow their evolution forward and  
197 backward in time, thus simulating the subjects' evolution from their original healthy condition to  
198 their current pathological state. We obtain z-scores trajectories summarizing the overall progression  
199 of AD, and from which we can estimate the long-term evolution of the corresponding clinical and  
200 imaging measurements.

201

---

202 **Disease severity quantification.** Relying on the reference trajectory estimated for the four z-  
203 scores summarizing AD evolution, we can subsequently perform individual disease staging. Based  
204 on the multi-modal imaging and clinical data of a given subject collected at any visit, we compute  
205 z-scores for each type of marker, and find the time-point  $\tau$  jointly minimizing the distance between  
206 the individual z-scores and the reference trajectory. In the rest of the paper, we will refer to this  
207 time-point as the disease severity. The estimated disease severity  $\tau$  locates a subject on the reference  
208 trajectory, thus quantifying its pathological state. It is also important to note that the disease severity  
209 can still be estimated even in the case of missing data, by only computing the z-scores of the  
210 available measures of the observed subject. We provide mathematical details on the disease severity  
211 estimation in Supplementary Section 1.

## 212 **2.4 Estimated model**

213 The parameters of the resulting model, presented in (Abi Nader et al. 2021), were estimated through  
214 the analysis of multi-modal longitudinal data from the ADNI cohort. The clinical scores consisted  
215 in the Clinical Dementia Rating Scale Sum of Boxes (CDRSB), Alzheimer’s Disease Assessment  
216 Scale (ADAS11), Functional Assessment Questionnaire (FAQ), Rey Auditory Verbal Learning Test  
217 (RAVLT) learning, RAVLT immediate, RAVLT forgetting and MMSE. Regional gray matter density,  
218 amyloid load and glucose metabolism were derived following the procedure detailed in Section 2.2.  
219 No tau-PET data was included in the model. Baseline socio-demographic information and summary  
220 statistics about clinical and imaging data for the subjects from the ADNI cohort are provided in Table  
221 1. The disease progression previously estimated by SimulAD on the ADNI database is illustrated  
222 in Supplementary Figure 2 in which we show the evolution of the z-scores and their associated  
223 imaging and clinical measures.

---

## 224 **2.5 Evaluation strategy**

225 We considered the model of evolution estimated by SimulAD on the ADNI cohort as the reference  
226 progression for AD. We evaluated SimulAD by conducting a series of experiments on both ADNI and  
227 GMC cohorts which aimed at demonstrating respectively the known-groups validity, the concurrent  
228 validity and the reliability of the model.

229

230 **Known-groups validity.** We evaluated how the estimated disease severity discriminates sub-  
231 jects across clinical groups within each cohort. It is expected that values of disease severity should  
232 increase along with the severity of the clinical status. We further assessed the group-wise consistency  
233 of the disease severity distribution, by comparing its values for similar clinical groups between  
234 ADNI and GMC cohorts. Differences between groups were assessed using Student's *t*-test and  
235 Cohen's *d* effect size.

236

237 **Concurrent validity.** We assessed the extent to which the estimated disease severity corre-  
238 lates with validated clinical and imaging assessments. For each subject MMSE score was available  
239 and imaging-biomarkers assessment as follows: hippocampal atrophy evaluated with Freesurfer,  
240 glucose metabolism and amyloid burden computed by extracting standardized uptake value ratio in a  
241 composite mask of regions of interest (MetaROI approach (Jagust et al. 2009; Landau, Harvey, et al.  
242 2010; Landau, Mintun, et al. 2012)). In the case of GMC, we also had 50 subjects who underwent  
243 a tau-PET scan. We compared their estimated disease severity with respect to their tau burden  
244 computed in a composite mask of relevant regions (MetaROI approach (Jack, Wiste, et al. 2017)).  
245 Correlation between the estimated disease severity and the different variables was assessed using  
246 Spearman rank correlation ( $\rho$ ).

---

247

248 **Reliability.** The two previous experiments aimed to quantify the validity of SimulAD in terms of  
249 disease staging based on the reference progression previously simulated on the ADNI database. We  
250 verified the consistency and robustness of the dynamics estimated by SimulAD by simulating the  
251 evolution of clinical and imaging-derived markers based on the GMC data. To this end, we applied  
252 the procedure described in Section 2.3 on the AD dementia subjects from the GMC cohort, thus  
253 providing us with a new model of progression for clinical and imaging measurements, as well as  
254 specific z-scores trajectories personalized to the GMC cohort. We compared the disease progression  
255 models obtained on ADNI and GMC cohorts by computing the average error between their z-scores  
256 trajectories over time. In addition to comparing the disease progression models obtained on both  
257 cohorts, we proposed to demonstrate the reliability of SimulAD by evaluating its prediction abilities  
258 at the individual level. To do so, we simulated the evolution of the MMSE for the subjects from the  
259 GMC cohort and compared the estimated values with the MMSE assessed by the physicians during  
260 the follow-up clinical visits.

## 261 **2.6 Data adjustment**

262 In order to implement the aforementioned assessment strategy, a number of additional pre-processing  
263 steps had to be carried out.

264

265 **Missing measures imputation.** We recall that SimulAD relies on 7 neuropsychological tests  
266 (CDRSB, MMSE, ADAS11, FAQ, RAVLT learning, RAVLT immediate and RAVLT forgetting)  
267 to compute the score  $z^{cli}$ , and that the only common clinical test between the ADNI and GMC  
268 cohorts is the MMSE. However, relying only on the MMSE would bias the computation of  $z^{cli}$   
269 for the subjects from the GMC cohort. To overcome this issue, we imputed the 6 missing clinical

---

270 scores for all the subjects from the GMC cohort. Imputation was carried out through k-neighbors  
271 regressions trained on the ADNI database to predict each clinical score based on the MMSE and the  
272 measures of regional grey matter volume. The trained models were subsequently applied on the  
273 GMC cohort to estimate the associated clinical scores. We performed a 10-fold cross validation on  
274 the ADNI database to evaluate the prediction performances of the models. We show in Table 2 the  
275 average and the 95% confidence interval of the relative error between the ground truth and predicted  
276 score. The average relative error remains below 10% for the CDRSB, ADAS11, RAVLT immediate  
277 and FAQ, while not exceeding 20% for the RAVLT learning and RAVLT forgetting. We provide  
278 additional information about the estimation error in Supplementary Section 3.1.

279

280 **FDG data harmonization.** We can observe in Table 1 that in the GMC cohort 36 subjects  
281 out of 93 are missing an FDG-PET scan, thus preventing the computation of their score  $z^{met}$ . For  
282 these subjects their disease severity can therefore be estimated only based on three z-scores ( $z^{cli}$ ,  
283  $z^{atr}$ ,  $z^{cli}$ ), leading to potential bias and mis-estimation compared to the rest of the cohort. To prevent  
284 this issue we computed the regional FDG uptake for every subjects of the GMC cohort based on  
285 their corresponding early-phase (6 min) of amyloid-PET scan (Daerr et al. 2017). This was done by  
286 fitting a linear regression between the early-amyloid and FDG uptake of all the subjects for each  
287 brain region. Figure 1 illustrates the linear fit between early-amyloid and FDG regional uptake for  
288 three brain regions. We show similar relationships for additional brain regions in Supplementary  
289 Section 3.2.

290

291 **Tracer bias correction.** The amyloid-PET scan of 17 subjects from the GMC cohort was  
292 acquired using the flutemetamol tracer, while the amyloid-PET scans of the remaining subjects from  
293 both ADNI and GMC cohorts were acquired using florbetapir. To compensate the effect of the tracer

294 on the regional uptake for these subjects, we converted the extracted amyloid burden on the centiloid  
 295 scale and back to a florbetapir scale (Klunk et al. 2015; Battle et al. 2018; Navitsky et al. 2018). We  
 296 observe in Figure 2 that the correction reduces the variability of the amyloid uptake values for the  
 297 flutemetamol group, increasing the overlapping with the florbetapir one. Supplementary Section 3.3  
 298 provides histograms illustrating the same effect in other brain regions.

Table 2: Relative error between the ground truth and k-neighbors prediction of the different clinical scores in the ADNI cohort. Average values and 95% confidence interval. CDRSB: Clinic Dementia Rating Scale Sum of Boxes; ADAS11: Alzheimer’s Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test. ADNI: Alzheimer’s Disease Neuroimaging Initiative. CI: Confidence interval.

Score	CDRSB	ADAS11	RAVLT immediate	RAVLT learning	RAVLT forgetting	FAQ
Relative error (%)	5.2	6.0	7.5	14.2	17.8	9.1
95% CI	[1.4 ; 10.3]	[3.2 ; 10.6]	[5.2 ; 10.2]	[10.7 ; 18.8]	[10.3 ; 25.8]	[1.5 ; 18.8]

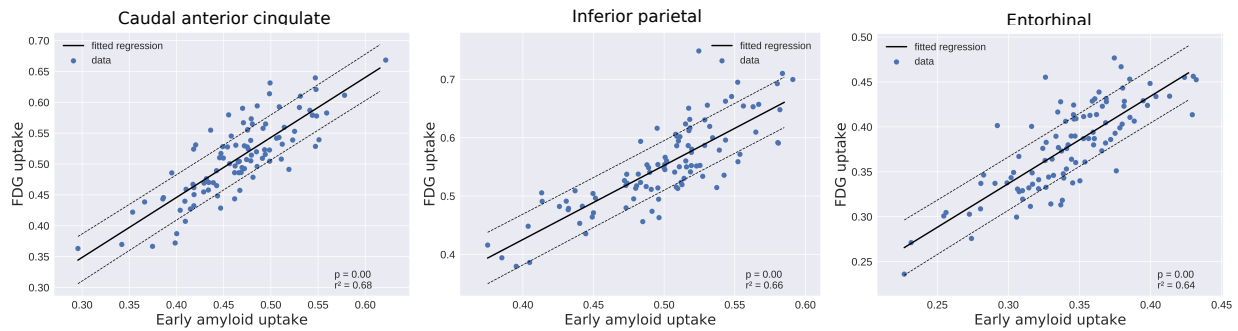


Figure 1: Scatter plot between the regional early-amyloid uptake and the corresponding FDG uptake for 57 patients of the GMC cohort. Solid black lines show the fitted linear model between regional FDG and early-amyloid. The dashed-lines represent 95% confidence interval. GMC: Geneva Memory Center FDG: (18)F-fluorodeoxyglucose.



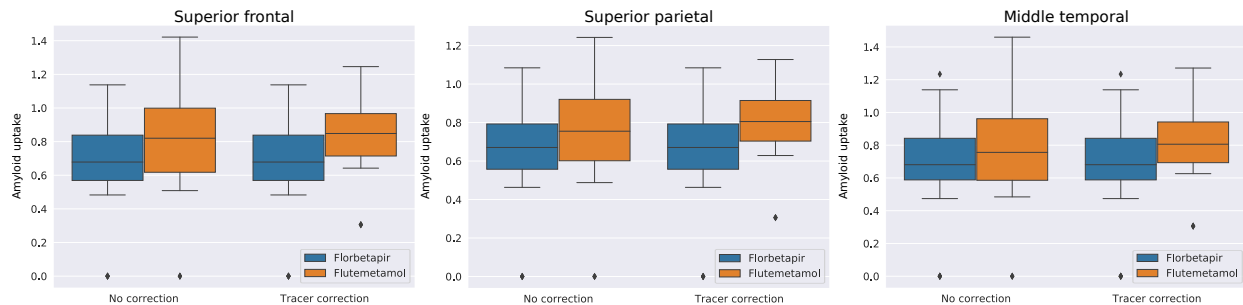


Figure 2: Distribution of the regional amyloid uptake depending on the tracer used during acquisition for the subjects from the GMC cohort. 76 amyloid-PET scans were acquired using florbetapir and 17 using flutemetamol. Tracer correction indicates that the regional amyloid uptake of subjects whose PET scan was acquired using flutemetamol was converted to a florbetapir scale. GMC: Geneva Memory Center; SUVR: Standardized Uptake Value Ratio; PET: Positron Emission Tomography.

## 299 3 Results

### 300 3.1 Known-groups validity

301 In this section, we considered the pathological progression previously estimated in (Abi Nader et al.  
 302 2021) by SimulAD as a reference trajectory, and computed the disease severity of the individuals  
 303 from both cohorts based on the procedure of Section 2.3. We show in Figure 3 the group-wise  
 304 distribution of the disease severity estimated by SimulAD for each subject in the ADNI and GMC  
 305 datasets. We observe that for both cohorts the disease severity increases when going from healthy  
 306 to pathological stages. The group-wise difference of disease severity across clinical groups is  
 307 statistically significant for each comparison (Student’s  $t$ -test  $p < 0.05$ ) except in the case of MCI  
 308 stable vs MCI converters for the GMC cohort (*cf.* Table 2a). We also notice rather large differences  
 309 between clinical groups ( $d > 0.7$ , *cf.* Table 2a) for both cohorts except in the case of NL stable vs  
 310 MCI stable for the ADNI cohort and MCI stable vs MCI converters for the GMC database. We also  
 311 evaluated the consistency of the disease severity by comparing its distribution for similar clinical

312 groups across cohorts. We observe in Table 2b that the estimated disease severity of similar clinical  
 313 groups is not significantly different ( $p > 0.05$ ), and differences between cohorts are rather small  
 314 ( $d \leq 0.1$ ), except in the case of MCI stable. We recall that, apart from the MMSE, the clinical scores  
 315 of the subjects from the GMC cohort were imputed based on the procedure detailed in Section 2.6.  
 316 We show in Supplementary Figure 7 that the estimation of the individual disease severity is robust  
 317 to this approximation.

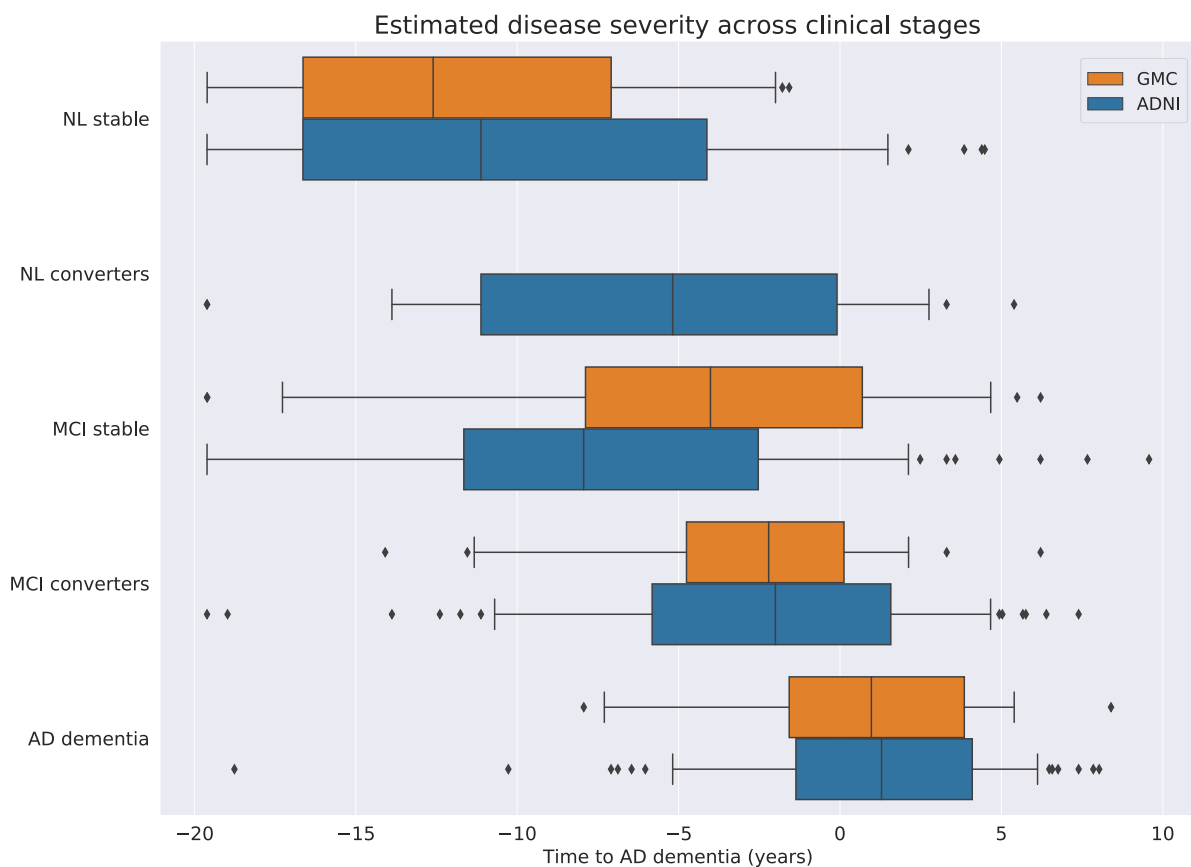


Figure 3: Distribution of the disease severity estimated by SimulAD across clinical stages for the ADNI and GMC cohorts relative to the estimated model of disease progression (Supplementary Figure 2). ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer’s dementia. Converters are NL patients progressing to MCI or AD dementia, or MCI individuals progressing to AD dementia.

Table 3: Comparison of the estimated disease severity distribution between clinical groups within each cohorts (a) and between similar clinical groups across cohorts (b); We report p-values of Student's *t*-test as well as the associated effect size (Cohen's *d*).

(a)

<b>Within cohorts disease severity comparison</b>											
	NL stable		NL stable		MCI stable		MCI stable		MCI converters		
	vs		vs		vs		vs		vs		
	NL converters		MCI stable		MCI converters		AD dementia		AD dementia		
Cohort	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC	
p-value	$7.5 \cdot 10^{-4}$	/	$1.8 \cdot 10^{-2}$	$1.0 \cdot 10^{-4}$	$1.2 \cdot 10^{-11}$	$4.4 \cdot 10^{-1}$	$1.1 \cdot 10^{-26}$	$3.8 \cdot 10^{-3}$	$3.1 \cdot 10^{-7}$	$1.6 \cdot 10^{-2}$	
Cohen's d	0.75	/	0.35	1.1	0.91	0.21	1.5	0.80	0.74	0.81	

(b)

<b>Between cohorts disease severity comparison</b>				
	NL stable	MCI stable	MCI converters	AD dementia
p-value	$8.2 \cdot 10^{-1}$	$7.0 \cdot 10^{-3}$	$9.5 \cdot 10^{-1}$	$7.3 \cdot 10^{-1}$
Cohen's d	0.03	0.57	0.10	0.09

---

## 318 **3.2 Concurrent validity**

319 We show in Figure 4 the progression of standard clinical and imaging markers with respect to the  
320 disease severity estimated by SimulAD for subjects from the ADNI and GMC databases. For both  
321 cohorts, the disease severity significantly correlates with the MMSE score (ADNI:  $\rho = -0.58$ ,  
322  $p < 0.01$ ; GMC:  $\rho = -0.55$ ,  $p < 0.01$ ). Regarding imaging-biomarkers, in both ADNI and GMC  
323 datasets the estimated disease severity correlates with hippocampal volume (ADNI:  $\rho = -0.57$ ,  
324  $p < 0.01$ ; GMC:  $\rho = -0.62$ ,  $p < 0.01$ ), glucose metabolism (ADNI:  $\rho = -0.80$ ,  $p < 0.01$ ; GMC:  
325  $\rho = -0.67$ ,  $p < 0.01$ ) and amyloid burden (ADNI:  $\rho = 0.44$ ,  $p < 0.01$ ; GMC:  $\rho = 0.31$ ,  $p < 0.01$ ).  
326 Since 50 individuals from the GMC cohort underwent a tau-PET scan, we also compare their  
327 estimated disease severity with their tau burden and show a significant correlation between them  
328 in Figure 4 ( $\rho = 0.62$ ,  $p < 0.01$ ). This latter correlation of the disease severity with a typical  
329 biomarker of AD that was not used for building the model supports the reliability of such a measure  
330 to summarize the overall severity of AD.

## 331 **3.3 Model reliability**

332 In the previous sections, we assessed the disease severity of the subjects based on the model of  
333 progression simulated on the ADNI cohort (Abi Nader et al. 2021). In order to demonstrate the  
334 reliability of the dynamics estimated by SimulAD, we simulated a new model of disease progression  
335 personalized to the GMC cohort. We show in Figure 5 the predicted evolution of imaging and  
336 clinical measurements based on this dataset. Similarly to what has been observed on the model  
337 previously trained on the ADNI database (Supplementary Figure 2), amyloid load increases and  
338 saturates early while following a uniform spatial pattern. Amyloid deposition is followed by a  
339 delayed process of neurodegeneration, more specifically a decrease of glucose metabolism and gray

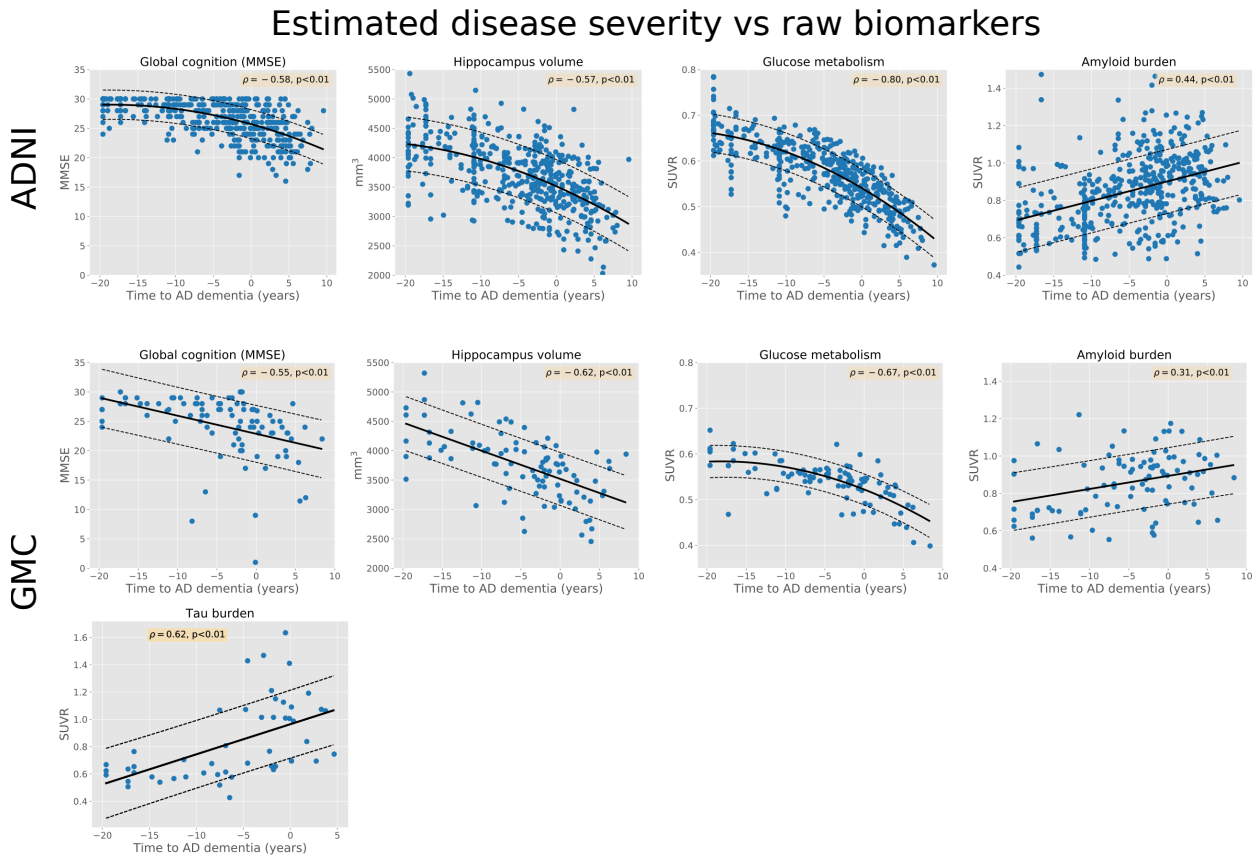


Figure 4: Relationships between the estimated disease severity and global cognition (MMSE), hippocampal volume, glucose metabolism, amyloid burden and tau burden (only for the GMC cohort). For each subject the estimated disease severity quantifies their position with respect to the model of disease progression (Supplementary Figure 2). The dashed-lines represent 95% confidence interval. ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center; SUVR: Standardized Uptake Value Ratio.

---

340 matter atrophy, mostly affecting temporal and parietal regions. Finally, clinical scores such as the  
341 MMSE or the CDRSB show a non-linear evolution accelerating during the latest stages of the disease.  
342

343 We also compared the models of progression obtained on the ADNI and GMC cohorts. Fig-  
344 ure 6 shows the evolution of the different z-scores depending on the cohort used to estimate the  
345 disease progression. These z-scores indicate the overall evolution of clinical scores, cerebral atrophy,  
346 amyloid deposition and glucose metabolism during AD. Given that the z-scores are not related to a  
347 physical unit but rather quantify the abnormality of a particular process, they were re-scaled between  
348 0 and 1 to illustrate the progression from healthy towards pathological stages. We observe that the  
349 four z-scores exhibit similar evolution patterns, whether they have been estimated on the ADNI or  
350 the GMC cohort. When averaged across time, the error between the z-scores of the two cohorts  
351 is of 6%, 7%, 8% and 9% for  $z^{cli}$ ,  $z^{atr}$ ,  $z^{met}$  and  $z^{amy}$  respectively. We provide in Supplementary  
352 Section 5 the evolution of the error between the evolutions of clinical scores and imaging regional  
353 measurements estimated based on the ADNI and the GMC cohort. When averaged over time, brain  
354 regions and clinical scores, the error is of 3%, 6%, 7% and 12% for MRI, FDG-PET, amyloid-PET  
355 derived regional measurements and clinical scores respectively. Finally, in spite of the fact that  
356 most of the clinical scores were imputed in the GMC cohort, Supplementary Section 4 shows that  
357 the resulting z-scores trajectories are robust to this estimation. Indeed, we observe that adding an  
358 additional error when imputing the clinical scores in the GMC cohort leads to rather small changes  
359 for the estimated z-scores trajectories compared to the results presented in Figure 6.

360 Finally, we applied SimulAD individually on all the patients from the GMC cohort to simulate  
361 their evolution over five years. We computed the error between their simulated MMSE and the one  
362 assessed by the physicians for the patients with available follow-up visits. The results are presented  
363 in Table 4, where we observe that the mean error ranges from 1.8 to 2.4 across clinical groups

364 highlighting the reliability of the model for estimating clinical outcomes during follow-up. For  
 365 illustrative purposes, we also provide in Supplementary Figure 10 the simulated evolution of the  
 366 MMSE, hippocampus volume, glucose metabolism and amyloid burden for four patients from the  
 367 GMC cohort.

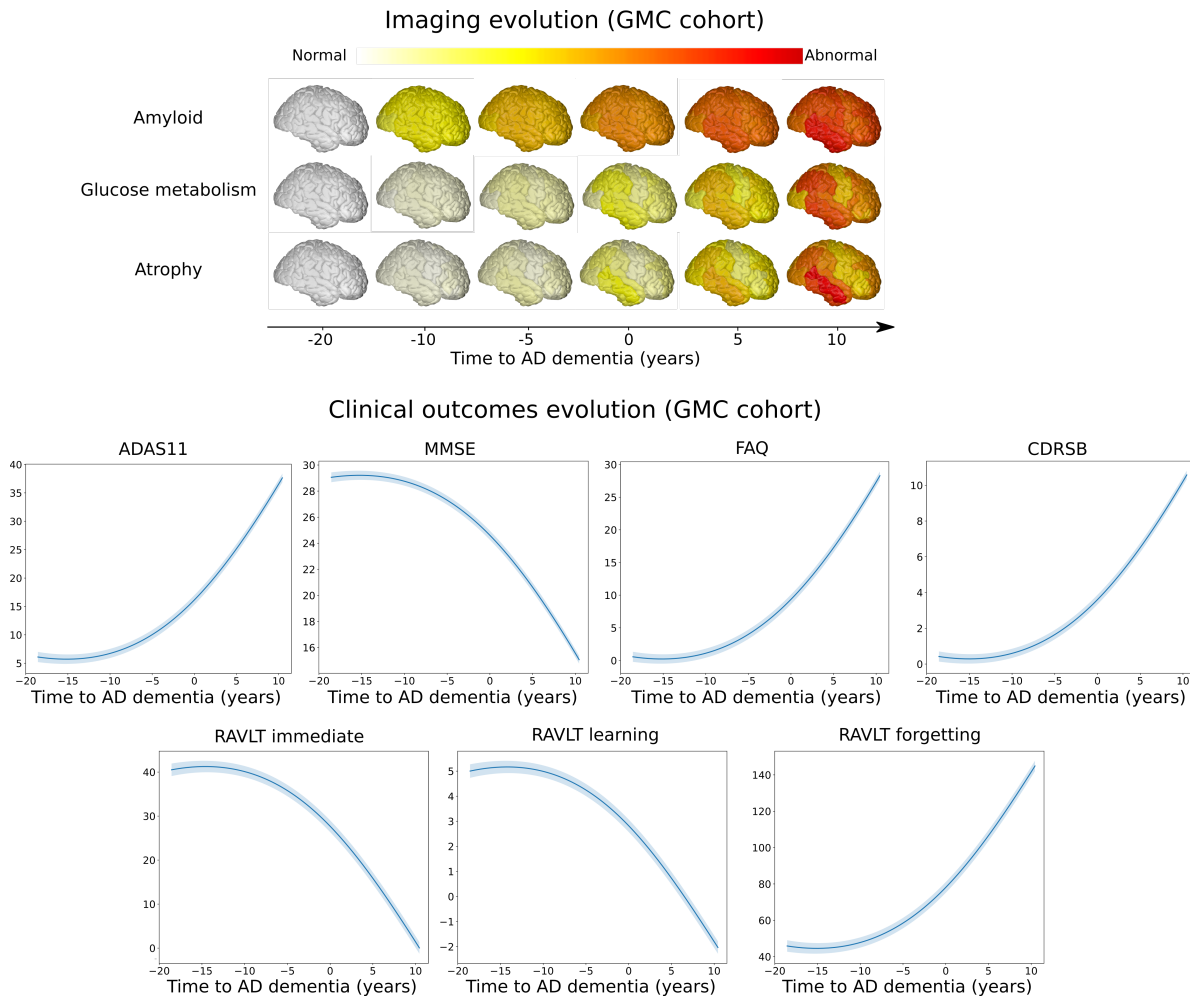


Figure 5: Simulated long-term evolution of cortical measurements for the different types of imaging markers and clinical scores based on the GMC cohort. Shaded areas represent the standard deviation of the average trajectory. GMC: Geneva Memory Center.

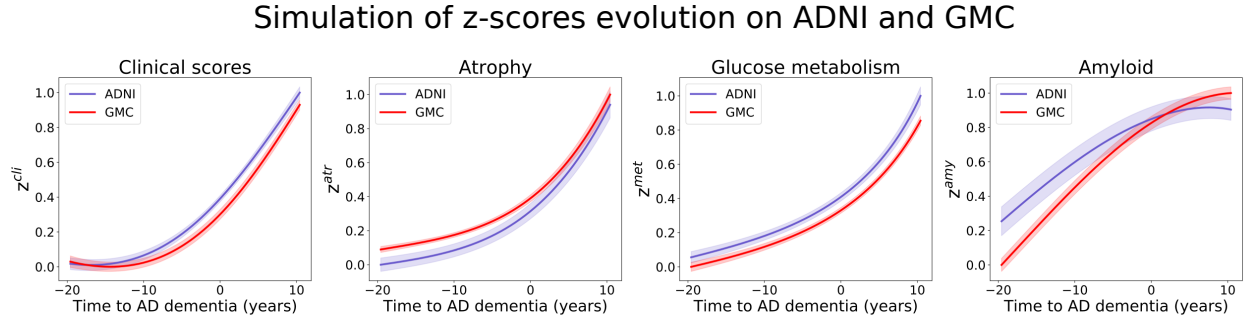


Figure 6: Estimated long-term dynamics depending on the cohort used for simulating the z-scores trajectories (time is relative to conversion to Alzheimer’s dementia). The z-scores have been re-scaled between 0 and 1 to illustrate the progression of each process from normal to pathological stages. ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center.

Table 4: For each clinical group: follow-up time in years (mean and standard deviation), total number of times the MMSE was assessed, and absolute error between the MMSE predicted by our method and the ground truth value from clinical evaluation (mean and standard deviation). NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer’s dementia. Converters are MCI individuals progressing to AD dementia.

Group	NL	MCI stable	MCI converters	AD dementia
Follow-up (years)	1.9 (1.2)	2.3 (1.5)	2.4 (1.7)	1.8 (1.4)
# MMSE assessments	48	63	59	31
Error	2.1 (1.5)	2.4 (1.9)	2.3 (2.0)	1.8 (1.6)



---

## 368 4 Discussion

369 In this study, we presented a thorough assessment of SimulAD on the independent GMC clinical  
370 cohort. The model was initially estimated based on the analysis of longitudinal imaging and clinical  
371 data from a subset of the ADNI database (Abi Nader et al. 2021). Due to a mismatch between cohorts  
372 in terms of missing measurements and data acquisition, we designed a pre-processing pipeline in  
373 order to enable the application of the model. We subsequently carried out a set of experiments to  
374 evaluate the validity of our approach and showed that the results obtained on the ADNI database  
375 were reproducible on the GMC cohort, outlining the robustness and generalization properties of  
376 SimulAD.

377

378 The present results underlined the feasibility of applying our model in a clinical context. We  
379 recall that throughout the experiments the disease severity was only assessed at baseline. Indeed, in  
380 the GMC cohort the subjects underwent the MMSE as well as an MRI, FDG-PET and amyloid-PET  
381 scan at baseline only. During the subsequent visits, only the MMSE was assessed in order to  
382 determine the clinical status of the subjects, however no imaging data was available. While the  
383 model is able to assess the disease severity even if many data modalities are missing, in this particular  
384 case the disease severity estimation would be entirely based on clinical information without any  
385 input from the imaging modalities. Given this context, our model wouldn't bring any additional  
386 insight about the pathological state of a subject compared to a battery of neuro-psychological tests.  
387 We showed that SimulAD was able to differentiate subjects across clinical groups for both cohorts  
388 (Figure 3, Table 3). Known-groups validity was established by assessing the similarity between  
389 cohorts of the estimated disease severity distribution for the NL stable, MCI converters and AD  
390 dementia groups (Table 2b). Moreover, the estimated disease severity correlated with clinical and

---

391 imaging assessments in both cohorts, and especially with tau burden on the GMC cohort. We recall  
392 that AD is defined by evidence of an abnormal load of both amyloid and tau (Jack, Bennett, et al.  
393 2018). Therefore, given that the model was estimated based on a subset of the ADNI cohort without  
394 accounting for tau-PET data, the correlation between the disease severity and tau burden on an  
395 external dataset is an appreciable demonstration of the concurrent validity of SimulAD. We also  
396 observed that independently from the cohort used to simulate the disease progression, the long-term  
397 evolution of the z-scores were similar between cohorts (Figure 6), thus outlining the reliability of  
398 the proposed model of AD progression. Finally, we showed that when applied on the GMC cohort,  
399 SimulAD was able to reliably predict the evolution of the MMSE of the subjects, highlighting its  
400 potential as a tool for individual monitoring of clinical outcomes. Overall, these results indicate  
401 general robustness of SimulAD when applied to independent cohorts characterized by different  
402 imaging acquisition protocols.

403

404 Some results highlighted in the proposed analysis deserve further discussion. There is no statistically  
405 significant difference between the estimated severity of the MCI stable and converters in the GMC  
406 cohort, while there is a significant difference of the estimated disease severity between the MCI  
407 stable of the ADNI and GMC cohorts. Concerning the first remark, it is important to note that  
408 among the 28 MCI stable subjects of the GMC cohort, a single visit only was available for 16 of  
409 them. It is therefore likely that this sub-group contains potential converters who may bias the MCI  
410 stable group with non-representative measurements. Regarding the second remark, we note in Figure  
411 3 that the median disease severity of the MCI stable group of the ADNI dataset is approximately  
412 of -8 years, while it is close to -4 years for the GMC cohort. This means that, according to our  
413 model, the pathological condition of the group of MCI stable from the GMC cohort is more severe  
414 than the one of the ADNI subset considered in the study. This is in agreement with the fact that

---

415 compared to the ADNI database, the MCI stable subjects from the GMC cohort exhibit a statistically  
416 significant lower MMSE score, lower volume of the hippocampus, lower glucose metabolism and  
417 higher amyloid burden (p-values < 0.05, Student's *t*-test). Moreover, we note that the distribution  
418 of the disease severity is conserved between both cohorts for NL stable, MCI converters and AD  
419 dementia groups. This result may point to the generalization capabilities of the disease severity  
420 estimation, which seems to be solely influenced by the clinical status. It also shows the ability of the  
421 model to identify specific clinical sub-types.

422

423 To our knowledge, the EBM and the DEBM are the only data-driven models of disease pro-  
424 gression which have been evaluated on several external cohorts (Archetti et al. 2019). The type  
425 of study here presented is therefore of relevant experimental value to enable the future application  
426 of disease progression models. We note that SimulAD presents certain advantages compared to  
427 standard event-based models, as it provides a finer-grained description of the disease for both spatial  
428 and temporal scales, in which regional changes affecting the brain are modelled on a continuous  
429 long-term time span. In addition, the method allows to personalize the evolution of clinical and  
430 imaging measurements for any patient or group of patients. Finally, the proposed mathematical  
431 formulation of AD progression as a system of ODEs allows to go beyond a simple description of  
432 the dynamics at stake during the disease evolution. In particular, we can simulate the impact of an  
433 anti-amyloid treatment on cognitive outcomes (Abi Nader et al. 2021).

434

435 As data-driven models are becoming more popular in healthcare thanks to their ability to leverage  
436 large scale clinical data, it is of utmost importance to facilitate their transfer from a research context to  
437 clinical practice. A growing number of regulatory institutions provided guidelines to help designing  
438 machine learning models that could be applied in clinical practice (Health 2019). The major problem

---

439 revolves around the generalization of the models beyond the dataset used to develop them, the main  
440 obstacle being their robustness to biases (Ghassemi et al. 2019). In this study, we developed a  
441 pipeline to mitigate the biases due to data heterogeneity and missing measurements. This work  
442 included data imputation for clinical data, and data standardization for PET imaging scans. Even  
443 though we carried out different tests to assess the reliability of our imputation strategy in Section  
444 2.6, it would be interesting in a future work to better evaluate the effect of the data imputation on  
445 the model performances. This could be done for instance by performing it on the ADNI cohort  
446 and re-assess the disease severity to observe if it led to substantial changes. Unfortunately, this  
447 cannot be done at the moment due to the choice of the method we used for approximating the  
448 missing neuro-psychological scores (k-neighbors), and also because no early-phase amyloid-PET  
449 was available to predict the corresponding FDG uptake on the sub-sample of the ADNI database  
450 used to train SimulAD. A more radical approach could be to simply discard all the missing data  
451 modalities, and re-train a new model with only the available data. We explored this possibility in  
452 Supplementary Section 7, and showed how it impacts the interpretability of the model as well as the  
453 disease severity estimation. Overall, even if SimulAD proved robust to these approximations, one  
454 of the main challenge highlighted in this study points to the complexity of data integration across  
455 cohorts and studies. In our particular case, neuropsychological tests could be standardized on a  
456 common scale, thus allowing to replace a score by another if they assess similar functions, while  
457 PET data could be systematically converted to the centiloid scale.

458

459 Finally, this study motivates further extensions of SimulAD that would foster its adoption in  
460 clinical practice. For instance, a limitation of our model is the underlying hypothesis that there  
461 exists a unique progression of AD which is common across individuals, while the disease is in  
462 fact highly heterogeneous. SimulAD could be extended in the future to account for multiple

---

463 risk factors, such as the presence of APOE4 (Kim et al. 2009), thus leading to a higher level of  
464 personalization of the predictions. Another avenue of improvement would be to account for a larger  
465 panel of biomarkers, such as tau (Pontecorvo et al. 2019), in order to better comprehend the disease  
466 progression. Currently, SimulAD is still a research software and future efforts should focus on the  
467 development of a user-friendly platform that could be deployed in clinical routine.

## 468 **5 Conclusion**

469 We presented a preliminary validation of SimulAD on the clinical cohort of the GMC. The results  
470 highlighted the reliability of the dynamics simulated by SimulAD for the disease key biomarkers,  
471 and showed encouraging performances in terms of disease staging on both cohorts. SimulAD is a  
472 promising modeling tool that may enable in the future the identification of subjects for enrollment in  
473 clinical trials, or the monitoring of the efficacy of disease modifying drugs.

## References

- 475 Abi Nader, Clément, Nicholas Ayache, Giovanni B Frisoni, Philippe Robert, and Marco Lorenzi  
476 (2021). “Simulating the outcome of amyloid treatments in Alzheimer’s Disease from multi-modal  
477 imaging and clinical data”. In: *Brain communications*.
- 478 Arbabshirani, Mohammad R., Sergey Plis, Jing Sui, and Vince D. Calhoun (2017). “Single subject  
479 prediction of brain disorders in neuroimaging: Promises and pitfalls”. In: *NeuroImage* 145, pp. 137–  
480 165. ISSN: 1053-8119. DOI: <https://doi.org/10.1016/j.neuroimage.2016.02.079>.
- 481 Archetti, D., S. Ingala, V. Venkatraghavan, V. Wottschel, A. L. Young, M. Bellio, E. E. Bron, S. Klein,  
482 F. Barkhof, D. C. Alexander, N. P. Oxtoby, G. B. Frisoni, and A. Redolfi (2019). “Multi-study  
483 validation of data-driven disease progression models to characterize evolution of biomarkers in  
484 Alzheimer’s disease”. In: *Neuroimage Clin* 24, p. 101954.
- 485 Battle, Mark R., Lovena Chedumbarum Pillay, Val J. Lowe, David Knopman, Bradley Kemp,  
486 Christopher C. Rowe, Vincent Doré, Victor L. Villemagne, and Christopher J. Buckley (2018).  
487 “Centiloid scaling for quantification of brain amyloid with [18F]flutemetamol using multiple  
488 processing methods”. In: *EJNMMI Research* 8.1, p. 107. ISSN: 2191-219X. DOI: 10.1186/s13550-  
489 018-0456-7.
- 490 Bilgel, M., J. L. Prince, D. F. Wong, S. M. Resnick, and B. M. Jernigan (2016). “A multivariate  
491 nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging”.  
492 In: *Neuroimage* 134, pp. 658–670.
- 493 Castro, Daniel C., Ian Walker, and Ben Glocker (2020). “Causality matters in medical imaging”. In:  
494 *Nature Communications* 11.1, p. 3673. ISSN: 2041-1723. DOI: 10.1038/s41467-020-17478-w.
- 495 Daerr, S., M. Brendel, C. Zach, E. Mille, D. Schilling, M. J. Zacherl, K. Bürger, A. Danek, O. Pogarell,  
496 A. Schildan, M. Patt, H. Barthel, O. Sabri, P. Bartenstein, and A. Rominger (2017). “Evaluation of  
497 early-phase [18F]-florbetaben PET acquisition in clinical routine cases”. In: *NeuroImage Clinical*  
498 14, pp. 77–86.
- 499 Davatzikos, Christos, Feng Xu, Yang An, Yong Fan, and Susan M. Resnick (2009). “Longitudinal  
500 progression of Alzheimer’s-like patterns of atrophy in normal older adults: the SPARE-AD index”.  
501 In: *Brain* 132.8, pp. 2026–2035. DOI: 10.1093/brain/awp091.
- 502 Desikan, R. S., F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner,  
503 A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany (2006). “An automated  
504 labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions  
505 of interest”. In: *NeuroImage* 31.3, pp. 968–980.
- 506 Falahati, Farshad, Eric Westman, and Andrew Simmons (2014). “Multivariate Data Analysis and  
507 Machine Learning in Alzheimer’s Disease with a Focus on Structural Magnetic Resonance  
508 Imaging”. In: *Journal of Alzheimer’s Disease* 41. 3, pp. 685–708. ISSN: 1875-8908. DOI: 10.  
509 3233/JAD-131928.
- 510 Fonteijn, H. M., M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill,  
511 S. J. Tabrizi, S. Ourselin, N. C. Fox, and D. C. Alexander (2012). “An event-based model for

---

512 disease progression and its application in familial Alzheimer’s disease and Huntington’s disease”.  
513 In: *NeuroImage* 60.3, pp. 1880–1889.

514 Gamberger, D., N. Lavrač, S. Srivatsa, R. E. Tanzi, and P. M. Doraiswamy (2017). “Identification of  
515 clusters of rapid and slow decliners among subjects at risk for Alzheimer’s disease”. In: *Sci Rep*  
516 7.1, p. 6763.

517 Ghassemi, Marzyeh, Tristan Naumann, Peter Schulam, Andrew L. Beam, Irene Y. Chen, and Rajesh  
518 Ranganath (2019). “Practical guidance on artificial intelligence for health-care data”. In: *The Lancet*  
519 *Digital Health* 1.4, e157–e159. ISSN: 2589-7500. DOI: 10.1016/S2589-7500(19)30084-6.

520 Greve, D. N., C. Svarer, P. M. Fisher, L. Feng, A. E. Hansen, W. Baare, B. Rosen, B. Fischl, and  
521 G. M. Knudsen (2014). “Cortical surface-based analysis reduces bias and variance in kinetic  
522 modeling of brain PET data”. In: *Neuroimage* 92, pp. 225–236.

523 Health, The Lancet Digital (2019). “Walking the tightrope of artificial intelligence guidelines in  
524 clinical practice”. In: *The Lancet Digital Health* 1.3, e100. ISSN: 2589-7500. DOI: 10.1016/S2589-  
525 7500(19)30063-9.

526 Jack, C. R., D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman,  
527 W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P. Rankin,  
528 C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, C. Elliott, E. Masliah, L. Ryan,  
529 and N. Silverberg (2018). “NIA-AA Research Framework: Toward a biological definition of  
530 Alzheimer’s disease”. In: *Alzheimers Dement* 14.4, pp. 535–562.

531 Jack, C. R., D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw,  
532 P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, and  
533 J. Q. Trojanowski (2013). “Tracking pathophysiological processes in Alzheimer’s disease: an  
534 updated hypothetical model of dynamic biomarkers”. In: *Lancet Neurol* 12.2, pp. 207–216.

535 Jack, C. R., H. J. Wiste, S. D. Weigand, T. M. Therneau, V. J. Lowe, D. S. Knopman, J. L. Gunter,  
536 M. L. Senjem, D. T. Jones, K. Kantarci, M. M. Machulda, M. M. Mielke, R. O. Roberts, P. Vemuri,  
537 D. A. Reyes, and R. C. Petersen (2017). “Defining imaging biomarker cut points for brain aging  
538 and Alzheimer’s disease”. In: *Alzheimers Dement* 13.3, pp. 205–216.

539 Jagust, W. J., S. M. Landau, L. M. Shaw, J. Q. Trojanowski, R. A. Koeppe, E. M. Reiman, N. L.  
540 Foster, R. C. Petersen, M. W. Weiner, J. C. Price, and C. A. Mathis (2009). “Relationships between  
541 biomarkers in aging and dementia”. In: *Neurology* 73.15, pp. 1193–1199.

542 Jedynak, B. M., A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, C. P. Jedynak,  
543 B. Caffo, and J. L. Prince (2012). “A computational neurodegenerative disease progression score:  
544 method and results with the Alzheimer’s disease Neuroimaging Initiative cohort”. In: *NeuroImage*  
545 63.3, pp. 1478–1486.

546 Khanal, Bishesh, Nicholas Ayache, and Xavier Pennec (2017). “Simulating Longitudinal Brain  
547 MRIs with known Volume Changes and Realistic Variations in Image Intensity”. In: *Frontiers in*  
548 *Neuroscience* 11. Article 132, p. 18. DOI: 10.3389/fnins.2017.00132.

549 Khanal, Bishesh, Marco Lorenzi, Nicholas Ayache, and Xavier Pennec (2016). “A biophysical model  
550 of brain deformation to simulate and analyze longitudinal MRIs of patients with Alzheimer’s  
551 disease”. In: *NeuroImage* 134, pp. 35–52. DOI: 10.1016/j.neuroimage.2016.03.061.

---

552 Kim, J., J. M. Basak, and D. M. Holtzman (2009). “The role of apolipoprotein E in Alzheimer’s  
553 disease”. In: *Neuron* 63.3, pp. 287–303.

554 Klöppel, Stefan, Cynthia M. Stonnington, Carlton Chu, Bogdan Draganski, Rachael I. Scahill,  
555 Jonathan D. Rohrer, Nick C. Fox, Jr Jack Clifford R., John Ashburner, and Richard S. J. Frackowiak  
556 (2008). “Automatic classification of MR scans in Alzheimer’s disease”. In: *Brain* 131.3, pp. 681–  
557 689. ISSN: 0006-8950. DOI: [10.1093/brain/awm319](https://doi.org/10.1093/brain/awm319).

558 Klunk, W. E., R. A. Koeppe, J. C. Price, T. L. Benzinger, M. D. Devous, W. J. Jagust, K. A. Johnson,  
559 C. A. Mathis, D. Minhas, M. J. Pontecorvo, C. C. Rowe, D. M. Skovronsky, and M. A. Mintun  
560 (2015). “The Centiloid Project: standardizing quantitative amyloid plaque estimation by PET”. In:  
561 *Alzheimers Dement* 11.1, pp. 1–15.

562 Koval, Igor, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie  
563 Allassonnière, and Stanley Durrleman (2018). “Spatiotemporal Propagation of the Cortical  
564 Atrophy: Population and Individual Patterns”. In: *Frontiers in Neurology* 9, p. 235. ISSN: 1664-  
565 2295. DOI: [10.3389/fneur.2018.00235](https://doi.org/10.3389/fneur.2018.00235).

566 Landau, S. M., D. Harvey, C. M. Madison, E. M. Reiman, N. L. Foster, P. S. Aisen, R. C. Petersen,  
567 L. M. Shaw, J. Q. Trojanowski, C. R. Jack, M. W. Weiner, and W. J. Jagust (2010). “Comparing  
568 predictors of conversion and decline in mild cognitive impairment”. In: *Neurology* 75.3, pp. 230–  
569 238.

570 Landau, S. M., M. A. Mintun, A. D. Joshi, R. A. Koeppe, R. C. Petersen, P. S. Aisen, M. W. Weiner,  
571 and W. J. Jagust (2012). “Amyloid deposition, hypometabolism, and longitudinal cognitive  
572 decline”. In: *Ann Neurol* 72.4, pp. 578–586.

573 Lorenzi, Marco, Maurizio Filippone, Giovanni B. Frisoni, Daniel C. Alexander, and Sebastien  
574 Ourselin (2017). “Probabilistic disease progression modeling to characterize diagnostic uncertainty:  
575 Application to staging and prediction in Alzheimer’s disease”. In: *NeuroImage*. ISSN: 1053-8119.  
576 DOI: <https://doi.org/10.1016/j.neuroimage.2017.08.059>.

577 Marinescu, R. V., A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, S. J. Crutch,  
578 and D. C. Alexander (2019). “DIVE: A spatiotemporal progression model of brain pathology in  
579 neurodegenerative disorders”. In: *NeuroImage* 192, pp. 166–177.

580 Mendelson, Alex F., Maria A. Zuluaga, Marco Lorenzi, Brian F. Hutton, and Sébastien Ourselin  
581 (2017). “Selection bias in the reported performances of AD classification pipelines”. In: *NeuroIm-  
582 age: Clinical* 14, pp. 400–416. ISSN: 2213-1582. DOI: [https://doi.org/10.1016/j.nicl.  
583 2016.12.018](https://doi.org/10.1016/j.nicl.2016.12.018).

584 Navitsky, Michael, Abhinay D. Joshi, Ian Kennedy, William E. Klunk, Christopher C. Rowe,  
585 Dean F. Wong, Michael J. Pontecorvo, Mark A. Mintun, and Michael D. Devous (2018).  
586 “Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to  
587 the Centiloid scale”. In: *Alzheimer’s & Dementia* 14.12, pp. 1565–1571. ISSN: 1552-5260. DOI:  
588 <https://doi.org/10.1016/j.jalz.2018.06.1353>.

589 Pontecorvo, M. J., M. D. Devous, I. Kennedy, M. Navitsky, M. Lu, N. Galante, S. Salloway,  
590 P. M. Doraiswamy, S. Southekal, A. K. Arora, A. McGeehan, N. C. Lim, H. Xiong, S. P.  
591 Trucchio, A. D. Joshi, S. Shcherbinin, B. Teske, A. S. Fleisher, and M. A. Mintun (2019). “A



---

592 multicentre longitudinal study of flortaucipir (18F) in normal ageing, mild cognitive impairment  
593 and Alzheimer’s disease dementia”. In: *Brain* 142.6, pp. 1723–1735.

594 Reuter, M., N. J. Schmansky, H. D. Rosas, and B. Fischl (2012). “Within-subject template estimation  
595 for unbiased longitudinal image analysis”. In: *NeuroImage* 61.4, pp. 1402–1418.

596 Sperling, R. A., C. R. Jack, and P. S. Aisen (2011). “Testing the right target and right drug at the  
597 right stage”. In: *Sci Transl Med* 3.111.

598 Venkatraghavan, Vikram, Esther E. Bron, Wiro J. Niessen, and Stefan Klein (2019). “Disease pro-  
599 gression timeline estimation for Alzheimer’s disease using discriminative event based modeling”.  
600 In: *NeuroImage* 186, pp. 518–532. ISSN: 1053-8119. DOI: [https://doi.org/10.1016/j.  
601 neuroimage.2018.11.024](https://doi.org/10.1016/j.neuroimage.2018.11.024).

602 Young, A. L., R. V. Marinescu, N. P. Oxtoby, M. Bocchetta, K. Yong, N. C. Firth, D. M. Cash,  
603 D. L. Thomas, K. M. Dick, J. Cardoso, J. van Swieten, B. Borroni, D. Galimberti, M. Masellis,  
604 M. C. Tartaglia, J. B. Rowe, C. Graff, F. Tagliavini, G. B. Frisoni, R. Laforce, E. Finger, A.  
605 de Mendonca, S. Sorbi, J. D. Warren, S. Crutch, N. C. Fox, S. Ourselin, J. M. Schott, J. D.  
606 Rohrer, and D. C. Alexander (2018). “Uncovering the heterogeneity and temporal complexity of  
607 neurodegenerative diseases with Subtype and Stage Inference”. In: *Nat Commun* 9.1, p. 4273.

---

## 608 **Acknowledgements**

609 **Funding** This work has been supported by the French government, through the UCA<sup>JEDI</sup> and 3IA  
610 Côte d’Azur Investments in the Future project managed by the National Research Agency (ref.n  
611 ANR-15-IDEX-01 and ANR-19-P3IA-0002), the grant AAP Santé 06 2017-260 DGA-DSH. The  
612 authors are grateful to the OPAL infrastructure from Université Côte d’Azur for providing resources  
613 and support.

614

615

616 **Data availability** Data for this study were collected at the Centre de la mémoire, Geneva Univer-  
617 sity and University Hospitals, thanks to funds from: Association Suisse pour la Recherche sur  
618 l’Alzheimer, Genève; Fondation Segré, Genève; Ivan Pictet, Genève; Fondazione Agusta, Lugano;  
619 Fondation Chmielewski, Genève; Fondation Privée des HUG, Genève, Velux Stiftung; Swiss  
620 National Science Foundation (projects n.320030\_182772, 320030\_169876 and 320030\_185028);  
621 Horizon 2020 (projects n. 667375); Human Brain Project; Innovative Medicines Initiatives (IMI  
622 contract n. 115736 and 115952).

623

## 1 Overview of SimulAD

We illustrate in Supplementary Figure 1 how SimulAD works. A weighted average of baseline measurements of amyloid deposition, glucose metabolism, atrophy and clinical scores allows to transform them in a corresponding z-score. A system of Ordinary Differential Equations (ODEs) links these four z-scores, thus providing us with the interaction rule allowing us to extrapolate the evolution of the z-scores in time, and to predict the corresponding imaging and clinical measures. In the next paragraphs, we provide details about the mathematical formulation of SimulAD.

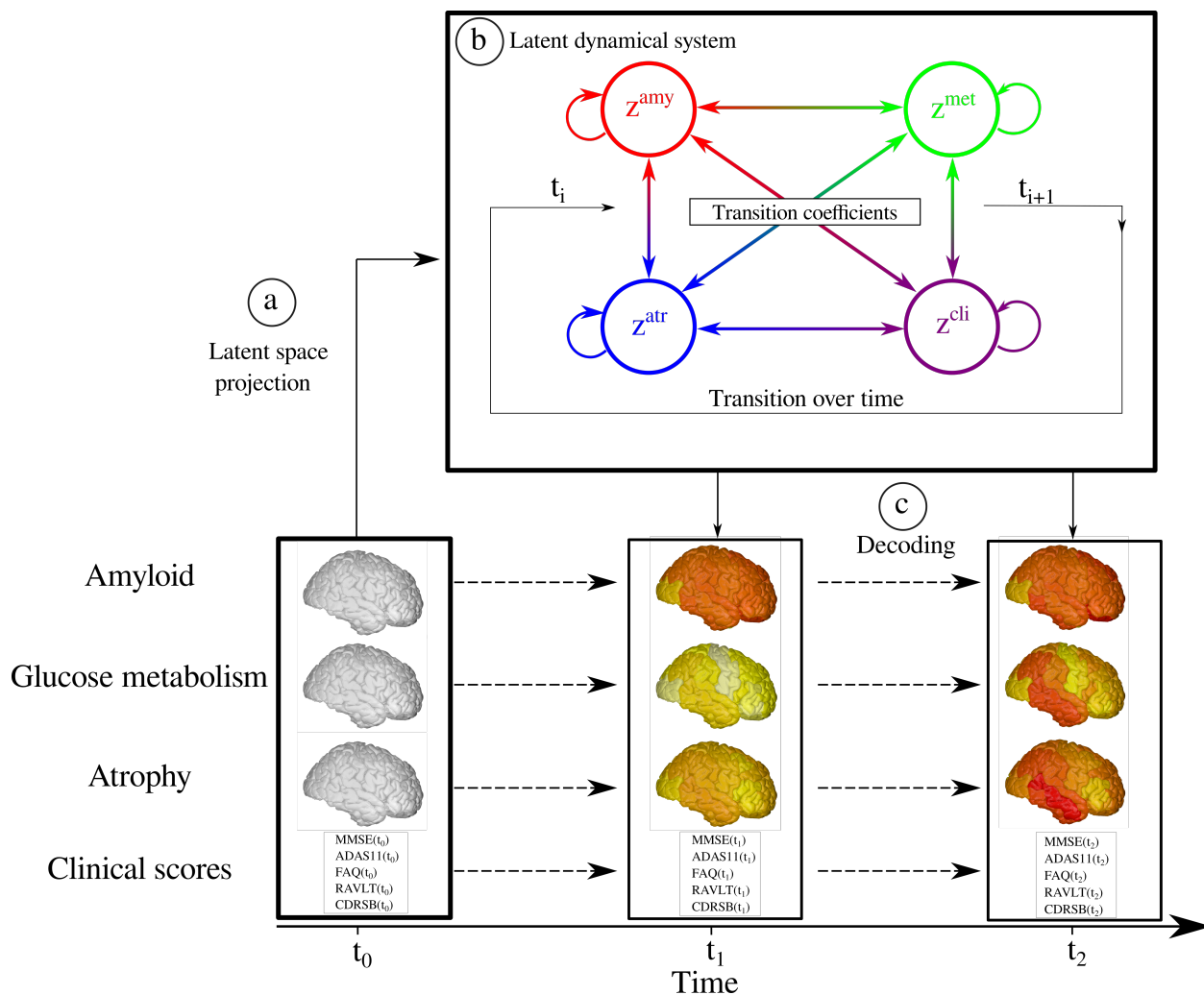
Measures from  $M$  data modalities at time  $t$  for a given subject  $i$  are stored in a matrix  $\mathbf{X}_i(t) = [\mathbf{x}_i^1(t), \mathbf{x}_i^2(t), \dots, \mathbf{x}_i^M(t)]^T$ . It is assumed that these measures are independently generated by a common latent representation  $\mathbf{z}_i(t)$  of the data such that:

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

$\sigma_m^2$  is measurement noise and  $\psi_m$  are the parameters of the function  $\mu_m$  that allows to compute the different measures of modality  $m$  for a subject based on its latent representation.  $\Lambda$  represents the interaction rule presented in the manuscript which allows to compute the latent variable at time  $t$ . It is defined as the solution of the following system of ODEs:

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

where the parameters of the system of ODEs  $\theta_{ODE}$  are the entries of the matrices  $\mathbf{W}$  and  $\mathbf{V}$ . This system of ODEs enforces a sigmoidal evolution for each type of modality  $m$ , while accounting for the relationship between the different modalities. Based on this system of ODEs and given baseline measures  $\mathbf{z}(0)$ , the variable  $\mathbf{z}(t)$  can be obtained by integration,  $\mathbf{z}(t) = \mathbf{z}(0) + \int_0^t g(\mathbf{z}(x), \theta_{ODE}) dx$ . This mathematical operation is the interaction rule used to compute the long-term trajectories of the z-scores.



Supplementary Figure 1: Overview of SimulAD. a) Measures of amyloid deposition, glucose metabolism, atrophy and clinical scores are transformed in a corresponding z-score  $z^{amy}$ ,  $z^{met}$ ,  $z^{atr}$ ,  $z^{cli}$ . 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 z-scores and the estimated dynamics, the follow-up measurements can be reconstructed to match the observed data.

---

Parameters' estimation is made by resorting to stochastic variational inference which is commonly used for this type of generative model. This gives us the following mathematical function that is maximized by gradient descent:

$$\mathcal{E} = \sum_m \mathbb{E}_{q(z|\mathbf{X})} \left[ \log p(\mathbf{x}^m | \mathbf{z}, \theta_{ODE}, \sigma_m^2, \psi_m) \right] - \mathcal{D} \left[ q(z^m | \mathbf{x}^m) | p(z^m) \right], \quad (3)$$

636 where  $\mathcal{D}$  refers to the Kullback-Leibler (KL) divergence, and  $q$  is an approximated posterior  
 637 distribution that allows to sample values  $z^m$  for a given data modality  $m$  based on the raw measures  
 638  $\mathbf{x}^m$ .

639  
 640 Finally, we can estimate the individual disease severity of a given subject based on its raw  
 641 measures  $\mathbf{X}$  and the reference progression  $\mathbf{z}(t)$ . The individual disease severity  $\tau$  is defined as the  
 642 time-point jointly minimizing the distance between the raw measures of the subject and the reference  
 643 progression:

$$\begin{aligned} \tau &= \arg \min_t \|f(\mathbf{X}, \phi^1) - \mathbf{z}(t)\|_1 \\ &= \sum_m |f(\mathbf{x}^m, \phi^1) - z^m(t)|. \end{aligned} \quad (4)$$

## 644 2 Model of disease progression

645 Once the model is optimized based on the ADNI data, we can simulate the evolution of four processes  
 646 at stake during AD: cognitive and behavioural decline, gray matter atrophy, amyloid deposition and  
 647 glucose metabolism. This is illustrated in Supplementary Figure 2 Panel I, in which we show the  
 648 estimated evolution of the z-scores by relying on the measures from the AD dementia subjects of the  
 649 ADNI cohort. The horizontal axis represents the course of the disease in years, and the reference  
 650 point  $t=0$  corresponds to the time of conversion to AD dementia. Given that the z-scores are not  
 651 related to a physical unit but rather quantify the abnormality of a particular process, they were  
 652 re-scaled between 0 and 1 to illustrate the progression from healthy towards pathological stages.  
 653 Notably  $z^{amy}$ , which summarizes the amyloid deposition, shows a fast abnormal progression in the  
 654 earliest stages before reaching a plateau. It is followed by  $z^{met}$  and  $z^{atr}$  respectively representing  
 655 a global decrease of glucose metabolism and gray matter atrophy, and which evolve at a similar  
 656 pace. Finally, we observe that the global clinical state quantified by  $z^{cli}$  shows a strong non-linear  
 657 acceleration after conversion to AD dementia.

658  
 659 Based on the z-score trajectories of Supplementary Figure 2 Panel I, we can compute the evolution

660 of their associated clinical and regional imaging measurements shown in Supplementary Figure  
 661 2 Panel II. We note that the progression of the clinical and imaging markers is compatible with  
 662 what has been observed for the z-scores. The predicted evolution of the remaining clinical scores is  
 663 provided in Supplementary Figure 3.

### 664 **3 Data correction**

#### 665 **3.1 Clinical scores imputation**

666 To complete what was presented in Section 2.6 concerning the imputation of the neuropsychological  
 667 tests for the subjects of the GMC cohort, we show in Supplementary Table 1 the Krippendorff  
 668 coefficient for each clinical score. This coefficient compares the inter-rater agreement, which would  
 669 be here the agreement between the ground truth score and the estimated one. We observe that the  
 670 value 0.8 is contained within the confidence interval of the CDRSB, ADAS11, RAVLT immediate  
 671 and FAQ, which makes the associated k-neighbors models reliable predictors according to the  
 672 thresholds defined by Krippendorff. The value 0.67 is contained within the confidence interval of  
 673 the RAVLT learning and RAVLT forgetting, which makes the corresponding k-neighbors models  
 674 acceptable for drawing prudent conclusions. We also provide the distribution of the absolute  
 675 prediction error for all the scores across clinical groups in Supplementary Figure 4.

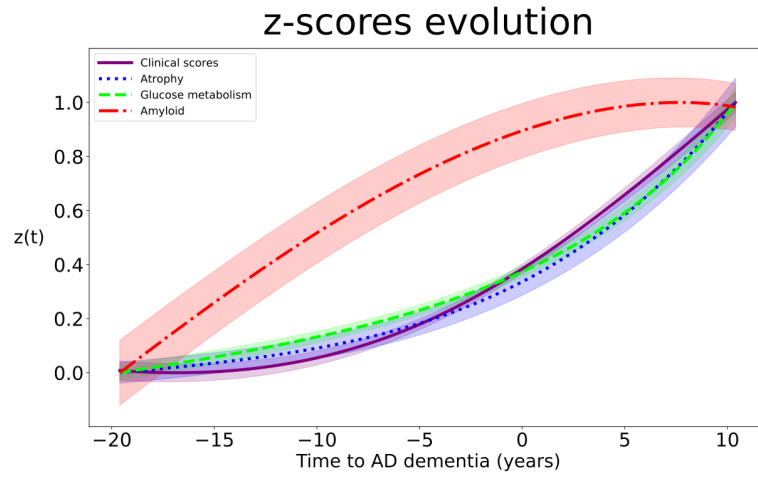
Supplementary Table 1: Krippendorff coefficient assessing the agreement between the ground truth and the k-neighbors prediction for all the clinical scores in the ADNI cohort. Average values and 95% confidence interval. CDRSB: Clinic Dementia Rating Scale Sum of Boxes; ADAS11: Alzheimer’s Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test. ADNI: Alzheimer’s Disease Neuroimaging Initiative. CI: Confidence interval.

Score	CDRSB	ADAS11	RAVLT immediate	RAVLT learning	RAVLT forgetting	FAQ
Krippendorff	0.77	0.77	0.80	0.61	0.69	0.78
95% CI	[0.69 ; 0.83]	[0.70 ; 0.83]	[0.74 ; 0.85]	[0.54 ; 0.68]	[0.62 ; 0.75]	[0.72 ; 0.84]

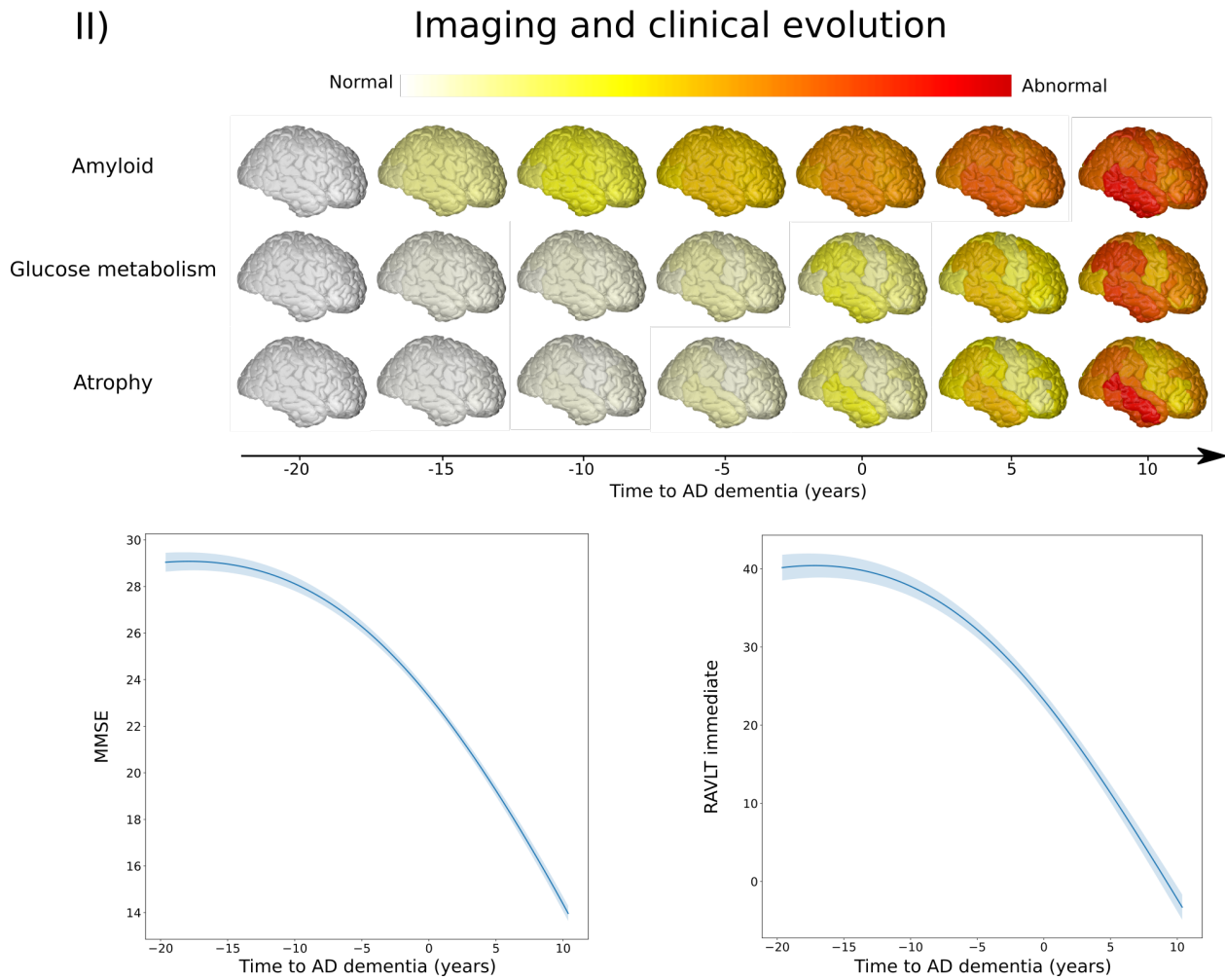
#### 676 **3.2 FDG prediction**

677 We show in Supplementary Figure 5 the fitted linear models between the FDG and early-AV45  
 678 uptake of 57 subjects from the GMC cohort for additional brain regions, as well as their raw values.

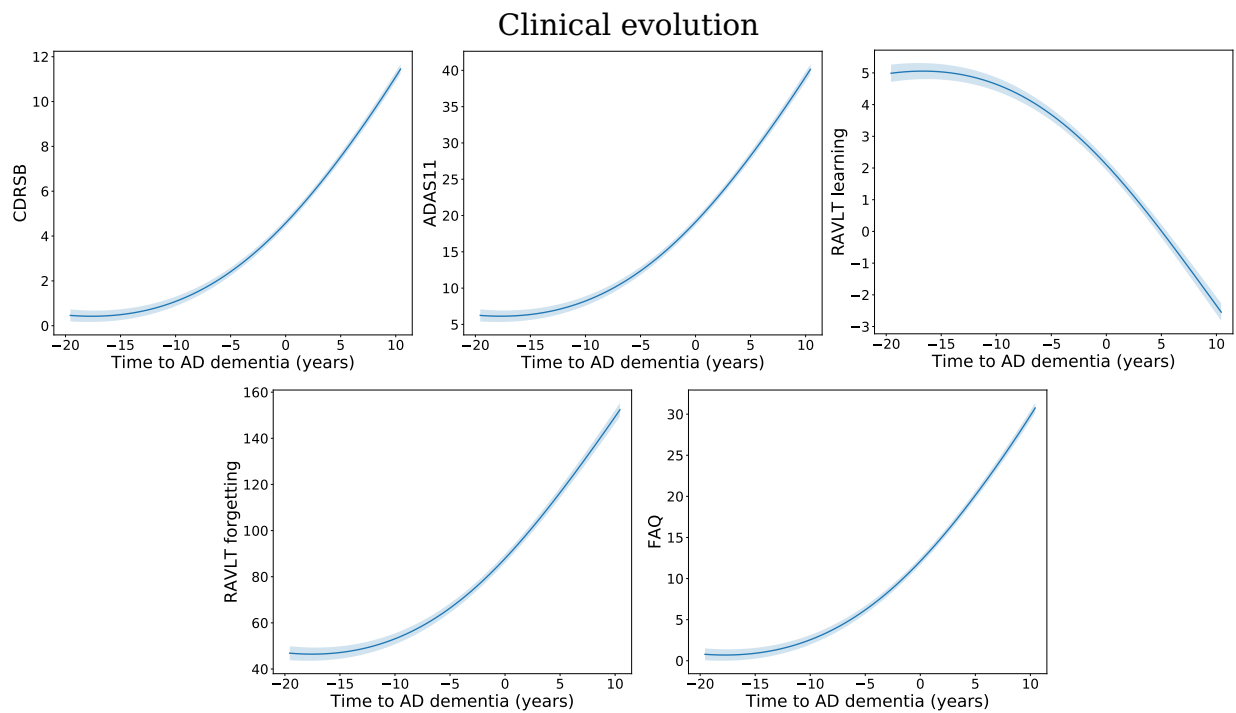
I)



II)

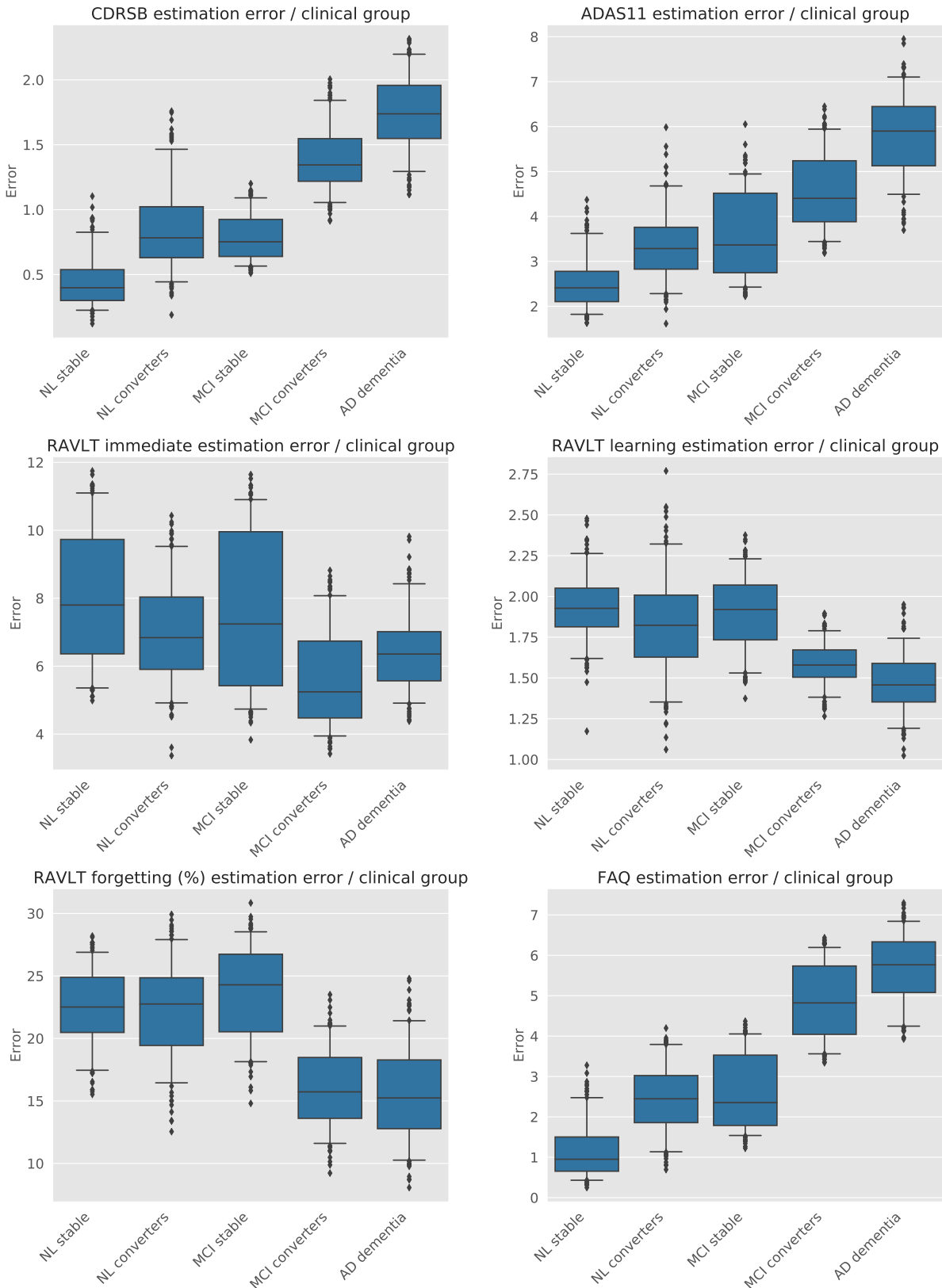


Supplementary Figure 2: Panel I: Estimated long-term dynamics of the z-scores (time is relative to conversion to Alzheimer’s dementia) based on the ADNI cohort. The z-scores were re-scaled between 0 and 1 to illustrate the progression of each process from normal to pathological stages. Shaded areas represent the standard deviation of the average trajectory. Panel II: Modelled long-term evolution of cortical measurements for the different types of imaging markers, and clinical scores based on the ADNI database. Shaded areas represent the standard deviation of the average trajectory.

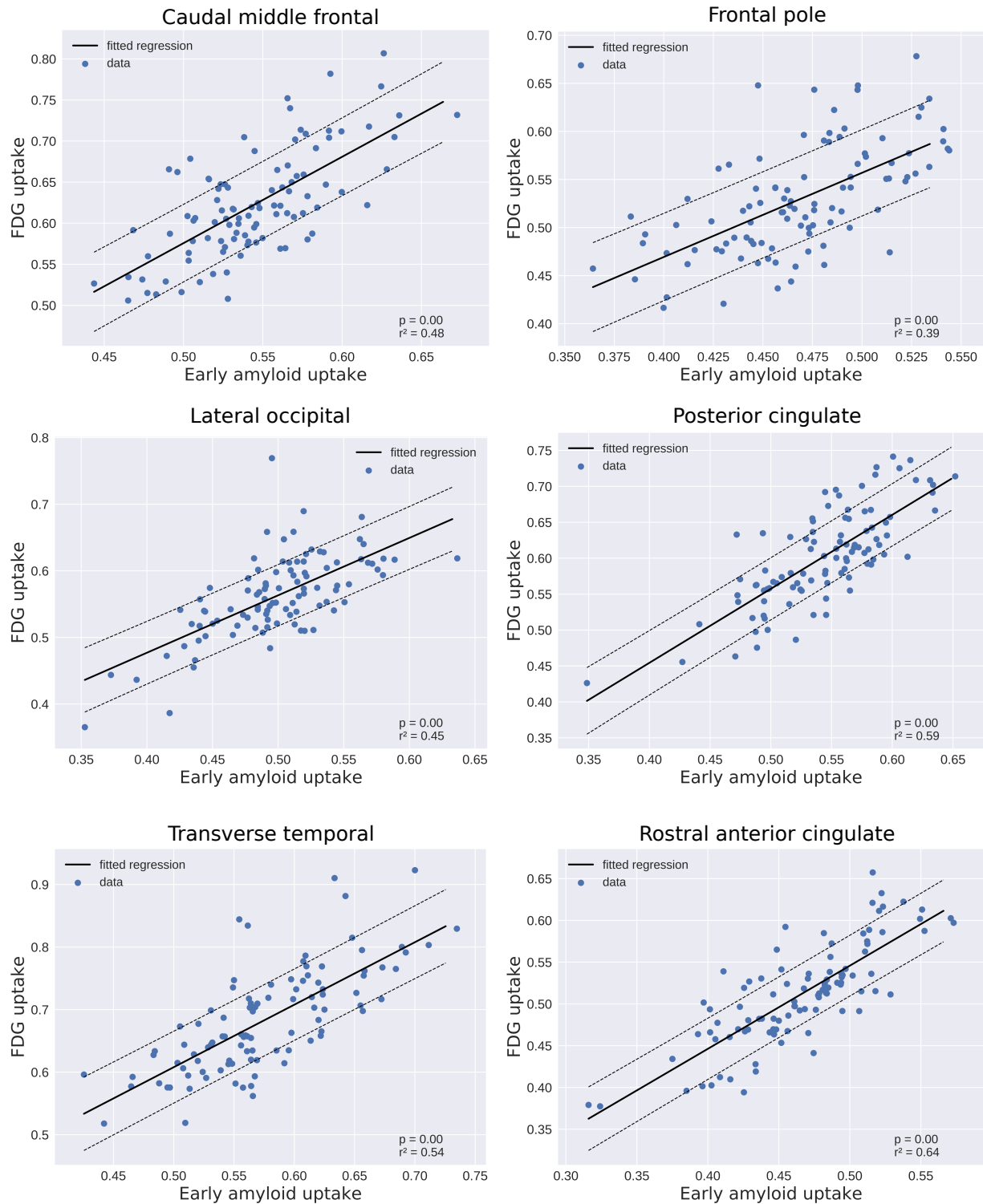


Supplementary Figure 3: Long-term modelled evolution of the missing clinical scores in Supplementary Figure 2. Shaded areas represent the standard deviation of the average trajectory. CDRSB: Clinic Dementia Rating Sum of Boxes; ADAS11: Alzheimer’s Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test.





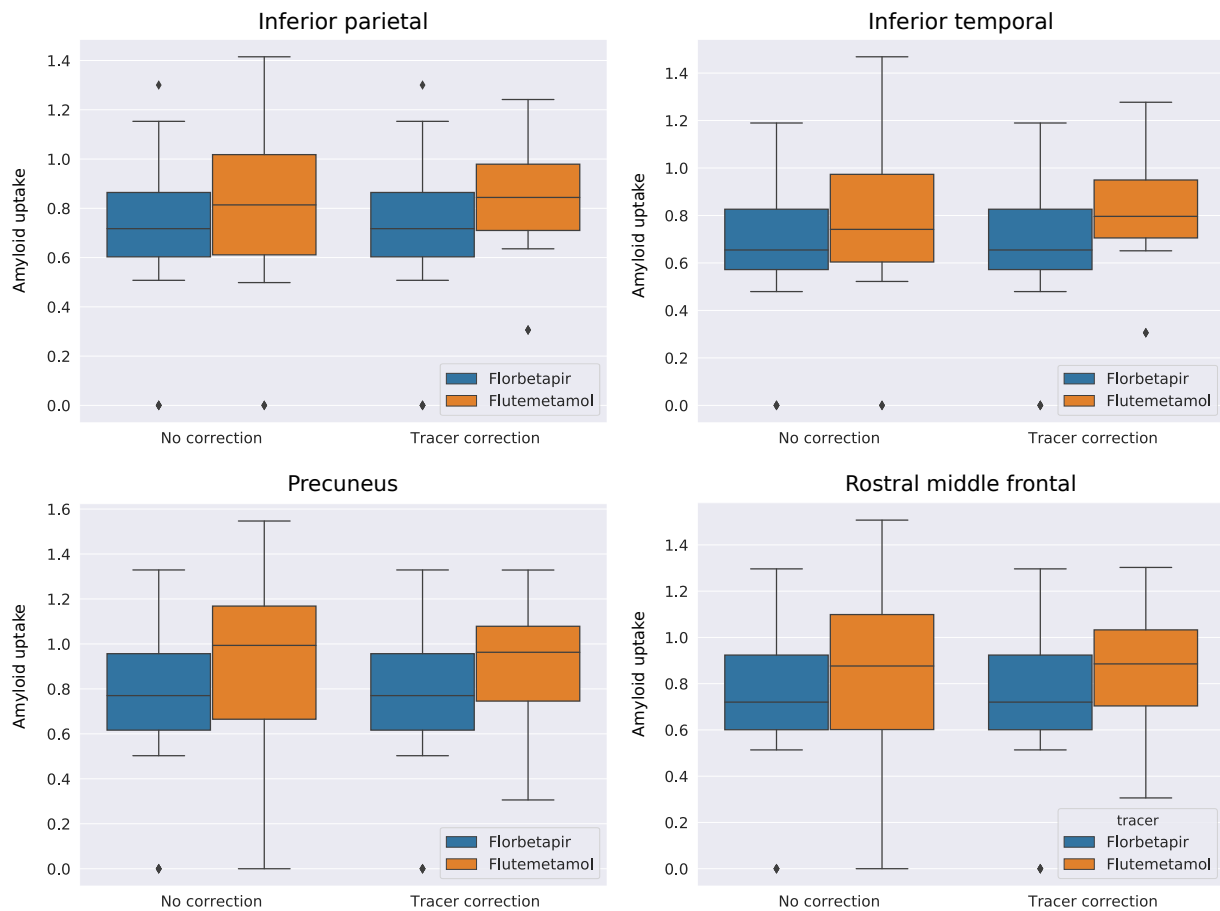
Supplementary Figure 4: Distribution of the absolute error between the ground truth and k-neighbors prediction for all the neuro-psychological tests across clinical groups. NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer’s dementia; CDRSB: Clinic Dementia Rating Sum of Boxes; ADAS11: Alzheimer’s Disease Assessment Scale; MMSE: Mini-Mental State Examination;



Supplementary Figure 5: Scatter plot between the regional early-amyloid uptake and the corresponding FDG uptake for 57 patients of the GMC cohort. Solid black lines show the fitted linear model between regional FDG and early-amyloid. The dashed-lines represent 95% confidence interval. GMC: Geneva Memory Center FDG: (18)F-fluorodeoxyglucose Positron Emission Tomography (PET) imaging.

679 **3.3 Tracer correction**

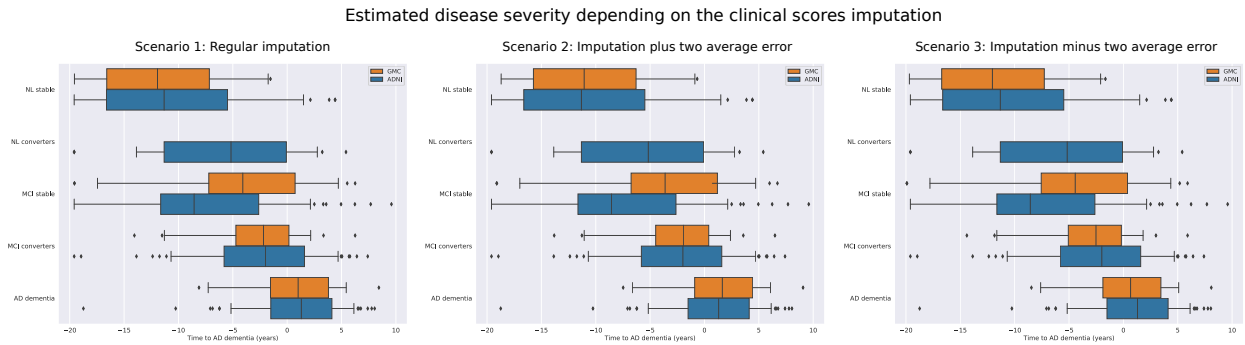
680 As mentioned in Section 2.6, we provide in Supplementary Figure 6 additional histograms showing  
681 the effect of the tracer correction in different brain regions. Consistently with what has been  
682 observed, we notice a reduction of the variability of the values for the flutemetamol group.



Supplementary Figure 6: Comparison of the distribution of the regional amyloid uptake before and after correction for the subjects from the GMC cohort, depending on the tracer used during acquisition. 76 Amyloid-PET scans were acquired using florbetapir and 17 using flutemetamol. GMC: Geneva Memory Center; SUVR: Standardized Uptake Value Ratio.

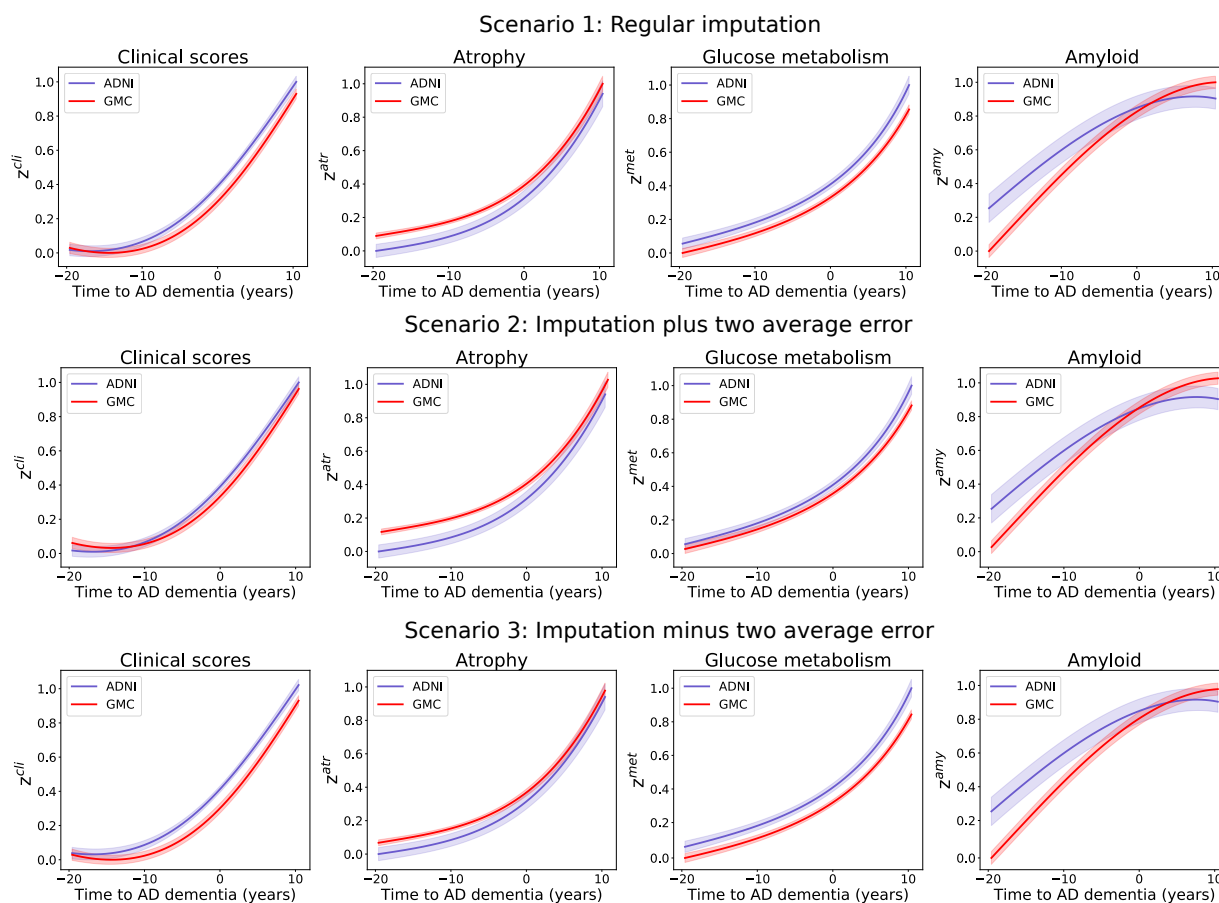
## 4 Robustness to the clinical scores prediction

684 In the case of the GMC cohort six clinical scores over seven were imputed based on the procedure  
 685 detailed in Section 2.6. These approximated neuropsychological assessments are used to compute  
 686 the z-scores  $z^{cli}$  of the individuals from the GMC cohort, thus affecting their estimated disease  
 687 severity shown in Figure 3, but also the progression of the z-scores for the GMC cohort in Figure 6.  
 688 We evaluate the impact of this approximation by computing the group-wise disease severity and the  
 689 progression of the z-scores in three scenarios: (i) clinical scores are imputed following the procedure  
 690 described in Section 2.6; (ii) clinical scores are imputed with an additional approximation by adding  
 691 twice their respective average prediction error; (iii) clinical scores are imputed with an additional  
 692 approximation by subtracting twice their average prediction error. We observe in Supplementary  
 693 Figures 7 and 8 that introducing this additional approximation error when imputing the clinical  
 694 scores leads to rather small changes compared with the results obtained with the regular imputation,  
 695 for both the group-wise disease severity and the z-scores progression. This shows that SimulAD is  
 696 robust to the imputation of the clinical scores for the GMC cohort, and that despite the induced  
 697 margin of error the estimated disease severity and z-scores progression are reliable.



Supplementary Figure 7: Effect of the clinical scores imputation on the estimated disease severity for the subjects from the GMC cohort. Scenario 1: Clinical scores are imputed following the procedure described in Section 2.6. Scenario 2: Clinical scores are imputed with an additional approximation by adding twice their respective average prediction error. Scenario 3: Clinical scores are imputed with an additional approximation by subtracting twice their average prediction error. ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer’s dementia. Converters are cognitively unimpaired and MCI subjects whose clinical diagnosis change during follow-up.

## Comparison of z-scores progression between ADNI and GMC cohorts depending on clinical scores imputation



Supplementary Figure 8: Effect of the clinical scores imputation on the long-term dynamics for the subjects from the GMC cohort. Shaded areas represent the standard deviation of the average trajectory. Scenario 1: Clinical scores are imputed following the procedure described in Supplementary Section 3.1. Scenario 2: Clinical scores are imputed with an additional approximation by adding twice their respective average prediction error. Scenario 3: Clinical scores are imputed with an additional approximation by subtracting twice their average prediction error. ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center.

---

## 698 **5 Comparison of the simulated evolution of clinical and imaging** 699 **measures based on ADNI and GMC cohorts**

700 The similarity between the two models of disease progression based on the ADNI and GMC cohorts  
701 (*cf.* Section 3.3) is confirmed in Supplementary Figure 9. In this figure, we show the temporal  
702 evolution of the error between the clinical and regional imaging measurements simulated based on  
703 the ADNI cohort and the ones based on the GMC cohort. We observe that for imaging markers  
704 the regional error remains below 10% for most regions across the different modalities, and reaches  
705 a maximum of approximately 20% in the case of amyloid deposition at the earliest stages of the  
706 disease. Concerning clinical scores, the error remains below 10% and 20% for MMSE and RAVLT  
707 immediate respectively. When averaged across time, brain regions and clinical scores, the error is of  
708 3%, 6%, 7% and 12% for MRI, FDG-PET, amyloid-PET data and clinical scores respectively.

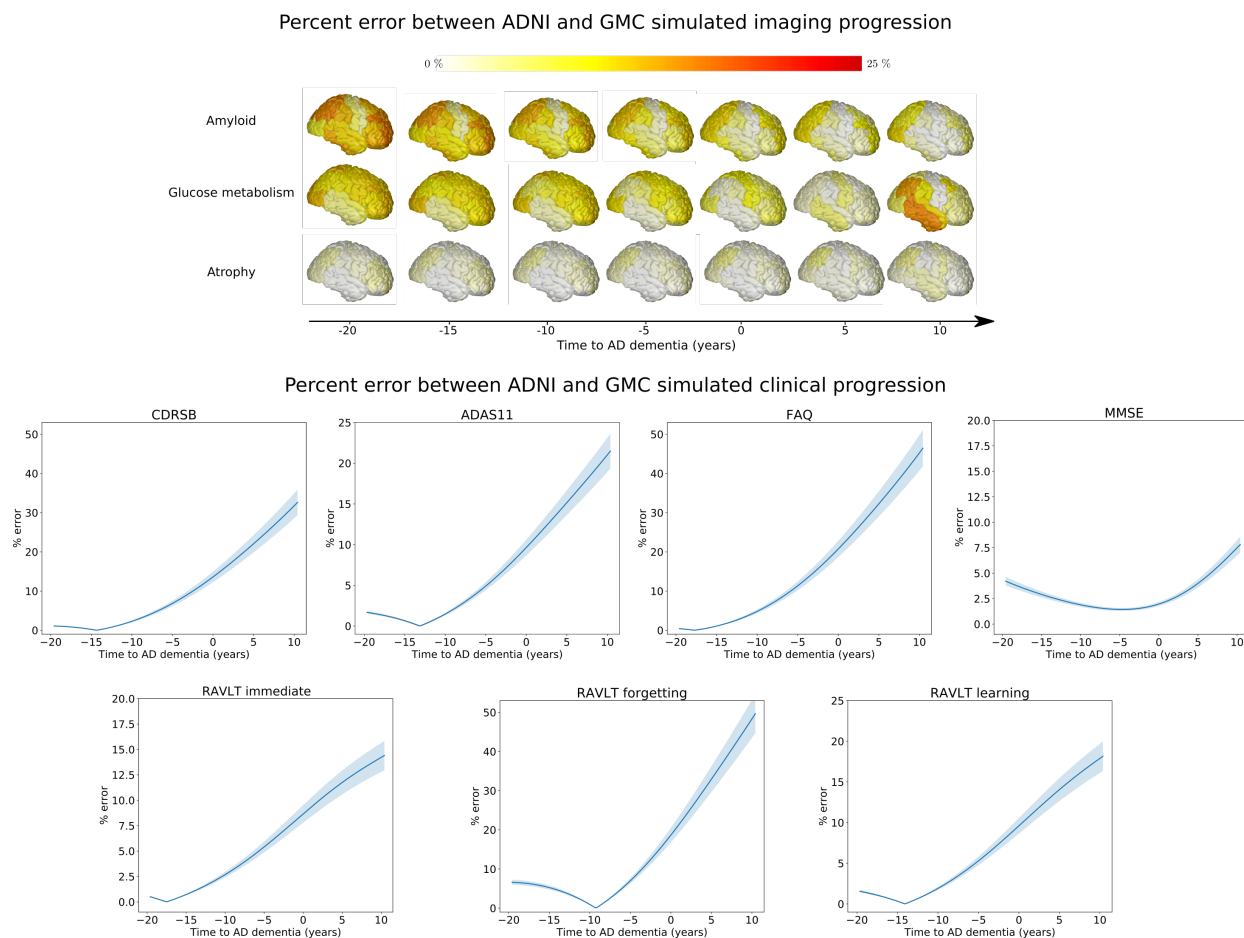
## 709 **6 Simulating individual trajectories**

710 In Supplementary Figure 10, we show the predicted evolution of the MMSE, hippocampus volume,  
711 glucose metabolism and amyloid burden for four subjects from the GMC cohort: one normal,  
712 one stable MCI, one MCI converter and one with AD dementia. We recall that the subjects from  
713 the GMC cohort underwent clinical assessments, MRI and PET scans at baseline only. During  
714 subsequent visits, the only available information were MMSE assessments. We observe that for  
715 these four patients, the model was able to accurately predict the MMSE. Concerning hippocampus  
716 volume, glucose metabolism and amyloid burden, we show the model prediction for the five next  
717 years but cannot compare it to the real values as this data was not available.

## 718 **7 Discarding FDG data and missing clinical data**

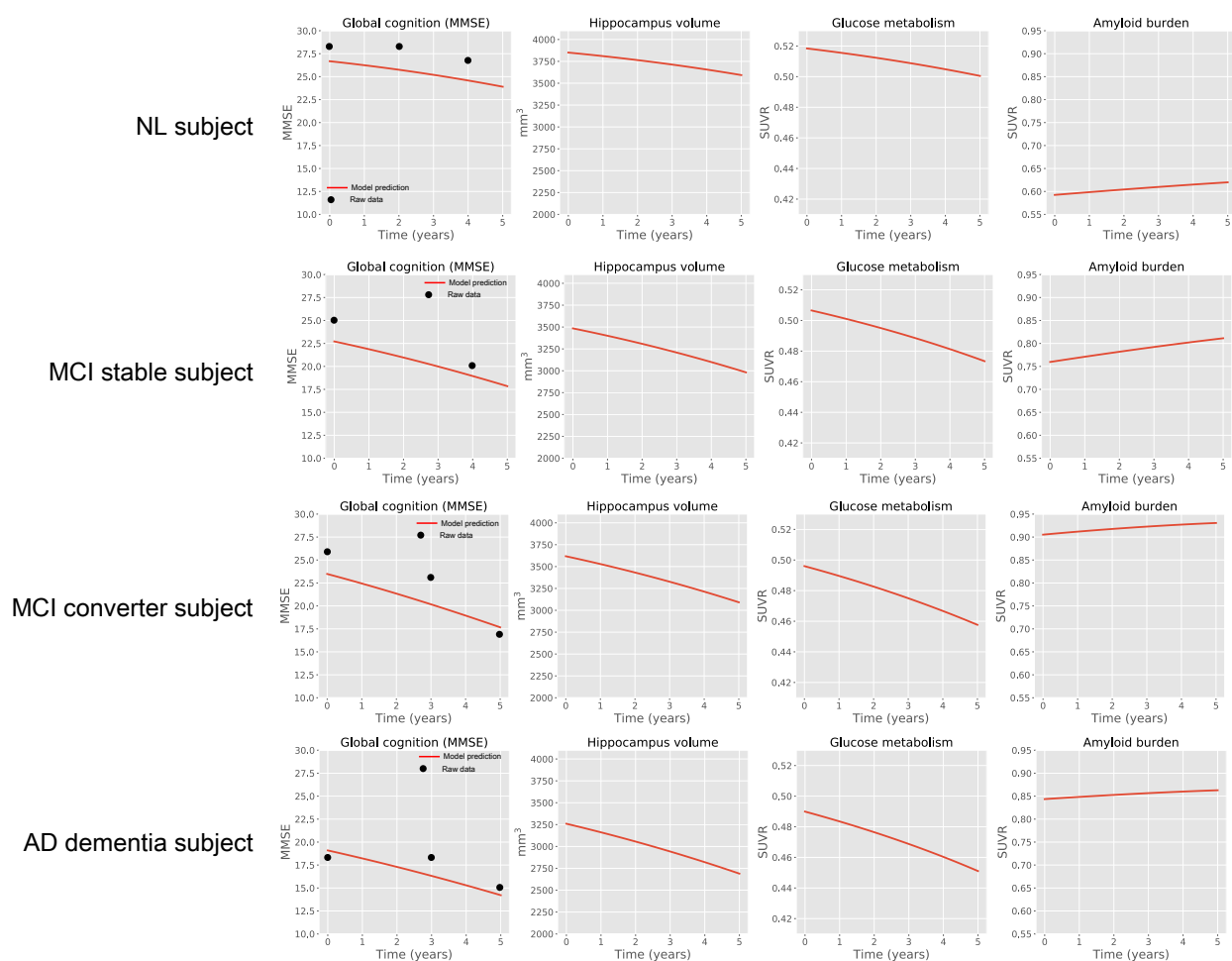
719 In this appendix, we propose to train a model on the ADNI cohort where the FDG data and the 6  
720 missing clinical scores are removed. We show in Figure 11 below the obtained trajectories for the  
721 three remaining z-scores. We observe that the model now spans approximately 20 years. Amyloid  
722 deposits and saturates first, followed by cerebral atrophy and cognitive decline. This sequence  
723 of events is in agreement with the hypothesis of the pathological cascade. However, compared  
724 to the original model we miss the information about glucose metabolism which is known to play  
725 an important role in the pathogenesis of AD. Moreover, the clinical state of patients is now only  
726 determined by their MMSE, while in the original model the clinical z-score was computed based  
727 on a total of seven neuro-psychological assessments, thus allowing to give a broader view of the  
728 clinical status of a subject.

729 Based on this new model, we estimated the disease severity of the subjects from the ADNI and GMC  
730 cohorts, as shown in Supplementary Figure 12. Similarly to Figure 3, the disease severity increases



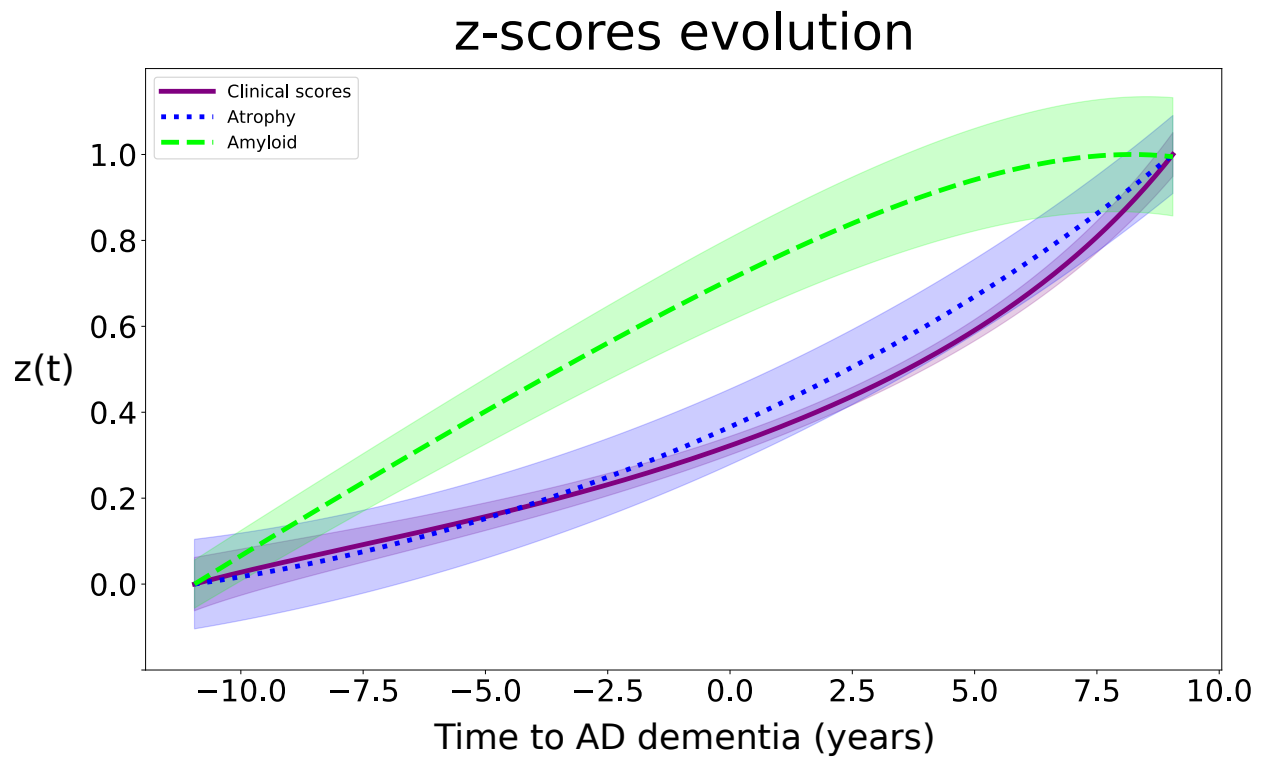
Supplementary Figure 9: Modelled long-term evolution of the error between the progressions simulated based on the ADNI and GMC cohorts in terms of cortical measurements and clinical scores. Shadowed areas represent the standard deviation of the average trajectory. ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center; CDRSB: Clinic Dementia Rating Sum of Boxes; ADAS11: Alzheimer’s Disease Assessment Scale; FAQ: Functional Assessment Questionnaire; RAVLT: Rey Auditory Verbal Learning Test; MMSE: Mini-Mental State Examination.

### Model prediction for subjects from GMC



Supplementary Figure 10: Prediction of MMSE, hippocampus volume, glucose metabolism and amyloid burden for four subjects from the GMC cohort. Red lines represent the model prediction and black dots raw values. GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer's dementia. Converters are NL patients progressing to MCI or AD dementia, or MCI individuals progressing to AD dementia.



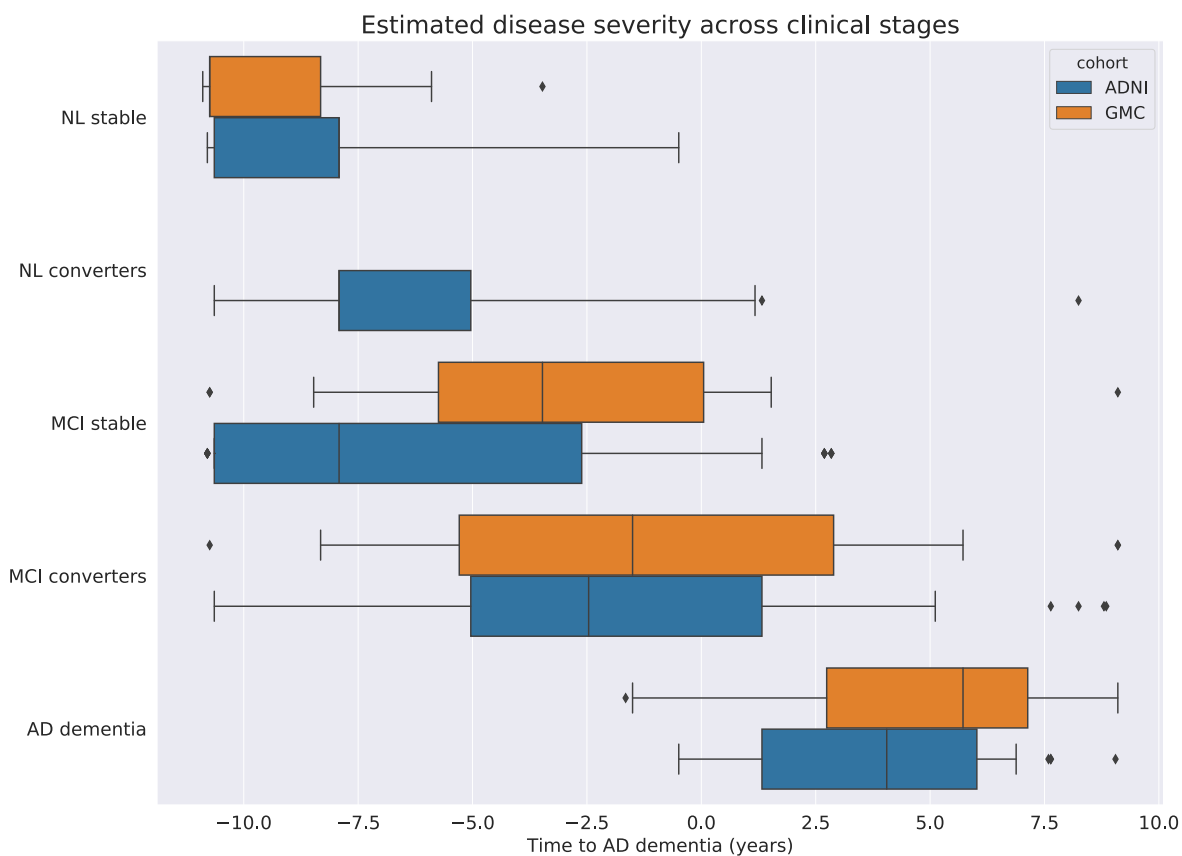


Supplementary Figure 11: Estimated long-term dynamics of the z-scores (time is relative to conversion to Alzheimer's dementia) based on the ADNI cohort. The z-scores were re-scaled between 0 and 1 to illustrate the progression of each process from normal to pathological stages. Shadowed areas represent the standard deviation of the average trajectory.

---

731 when going from healthy towards pathological stages. The group-wise difference of disease severity  
732 between clinical groups is statistically significant except in the case of MCI stable and converters for  
733 the GMC cohort (*cf.* Supplementary Table 2). We also observe rather large differences between  
734 clinical groups ( $d > 0.7$  *cf.* Supplementary Table 2) except for NL stable vs MCI stable for the ADNI  
735 cohort and MCI stable vs MCI converters for the GMC database, as shown in the manuscript where  
736 all the data modalities were included. Finally, if we look at the distribution of the disease severity  
737 for similar groups between cohorts, we notice that differences between cohorts are larger compared  
738 to the original experiment. Indeed, we see that the Cohen's  $d$  is approximately of 0.3, while before it  
739 was below 0.1 (Table 3) for NL stable, MCI converters and AD dementia, while in the case of the  
740 MCI stable it increased from 0.57 to 0.74.

741 Overall, this shows that even if we remove FDG data and the missing clinical scores, SimulAD is  
742 still able to learn a disease progression model that can accurately stage patients and differentiate  
743 between clinical groups. However, we obtain a less informative model which doesn't take into  
744 account the diversity of pathological processes affecting the brain during AD. We also noticed that  
745 the disease severity for similar groups across cohorts was not as consistent as in the original case.  
746 This might be due to the fact that despite the approximation due to the imputation of the FDG  
747 uptake and missing clinical scores, the inclusion of this data still allows a more global assessment  
748 of subjects. Indeed, the heterogeneity of the data on which our model relies is essential to fully  
749 appreciate the pathological state of a subject. Therefore, we would recommend to take advantage of  
750 the ease to include heterogeneous data in our model, and use SimulAD with as many data modalities  
751 as possible, even if this might require some pre-processing steps to impute missing data. However,  
752 in the case where imputation would be too cumbersome or impossible, we showed in this experiment  
753 that SimulAD would still provide a reliable estimation of the disease severity of patients when some  
754 data modalities are discarded.



Supplementary Figure 12: Distribution of the disease severity estimated by SimulAD across clinical stages for the ADNI and GMC cohorts relatively. The disease severity is estimated relatively to a progression model in which FDG data and the 6 missing clinical scores were discarded and no imputation step needs to be carried out. ADNI: Alzheimer’s Disease Neuroimaging Initiative; GMC: Geneva Memory Center; NL: cognitively healthy; MCI: mild cognitive impairment; AD: Alzheimer’s dementia. Converters are NL patients progressing to MCI or AD dementia, or MCI individuals progressing to AD dementia.

Supplementary Table 2: Comparison of the estimated disease severity distribution between clinical groups within each cohorts (a) and between similar clinical groups across cohorts (b); We report p-values of Student's *t*-test as well as the associated effect size (Cohen's *d*).

(a)

<b>Within cohorts disease severity comparison</b>										
	NL stable vs NL converters		NL stable vs MCI stable		MCI stable vs MCI converters		MCI stable vs AD dementia		MCI converters vs AD dementia	
Cohort	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC	ADNI	GMC
p-value	$1,1 \cdot 10^{-2}$	/	$2,5 \cdot 10^{-6}$	$1,1 \cdot 10^{-7}$	$5,4 \cdot 10^{-6}$	$1,9 \cdot 10^{-1}$	0.0	$3,3 \cdot 10^{-7}$	0.0	$3,3 \cdot 10^{-4}$
Cohen's d	0.62	/	0.65	1.73	0.75	0.38	2.60	1.81	1.52	1.21

(b)

<b>Between cohorts disease severity comparison</b>				
	NL stable	MCI stable	MCI converters	AD dementia
p-value	$1,1 \cdot 10^{-1}$	$4,4 \cdot 10^{-3}$	$2,9 \cdot 10^{-1}$	$3,5 \cdot 10^{-1}$
Cohen's d	0.32	0.74	0.30	0.33