



**HAL**  
open science

## Determining Clinical Disease Progression in Symptomatic Patients With CADASIL

Sofia Kaisaridi, Dominique Herve, Aude Jabouley, Sonia Reyes, Carla Machado, Stéphanie Guey, Abbas Taleb, Fanny Fernandes, Hugues Chabriat, Sophie Tezenas Du Montcel

► **To cite this version:**

Sofia Kaisaridi, Dominique Herve, Aude Jabouley, Sonia Reyes, Carla Machado, et al.. Determining Clinical Disease Progression in Symptomatic Patients With CADASIL. *Neurology*, 2025, 104 (1), 10.1212/WNL.000000000210193 . hal-04848392

**HAL Id: hal-04848392**

**<https://hal.science/hal-04848392v1>**

Submitted on 19 Dec 2024

**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.

## Determining clinical disease progression in symptomatic patients with CADASIL

### **Authors:**

Sofia Kaisaridi<sup>1</sup>; Dominique Herve<sup>2,3</sup>; Aude Jabouley<sup>2</sup>; Sonia Reyes<sup>2</sup>; Carla Machado<sup>2</sup>; Stéphanie Guey<sup>2,3</sup>; Abbas Taleb<sup>2</sup>; Fanny Fernandes<sup>2</sup>; Hugues Chabriat<sup>2,3</sup>; Sophie Tezenas Du Montcel<sup>1</sup>

1. ARAMIS, Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Groupe Hospitalier Sorbonne Université, Paris, France; 2. Centre de Référence pour les maladies vasculaires rares du cerveau et de l'œil (CERVCO) and Centre Neurovascular Translationnel (CNVT), AP-HP; 3. INSERM U1141 - FHU NeuroVasc, Université Paris Cité

**Corresponding Author:** Sophie Tezenas Du Montcel

**Group Authorship:** No. There is no study group involved in our research

**Group Name:**

### **Contributions:**

Sofia Kaisaridi: Drafting/revision of the manuscript for content, including medical writing for content; Study concept or design; Analysis or interpretation of data; Dominique Herve: Major role in the acquisition of data; Study concept or design; Aude Jabouley: Major role in the acquisition of data; Analysis or interpretation of data; Sonia Reyes: Major role in the acquisition of data; Analysis or interpretation of data; Carla Machado: Major role in the acquisition of data; Analysis or interpretation of data; Stéphanie Guey: Major role in the acquisition of data; Analysis or interpretation of data; Abbas Taleb: Major role in the acquisition of data; Analysis or interpretation of data; Fanny Fernandes: Major role in the acquisition of data; Analysis or interpretation of data; Hugues Chabriat: Drafting/revision of the manuscript for content, including medical writing for content; Major role in the acquisition of data; Study concept or design; Analysis or interpretation of data; Sophie Tezenas Du Montcel: Drafting/revision of the manuscript for content, including medical writing for content; Study concept or design; Analysis or interpretation of data;

**Abstract:** Background and Objectives: Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy (CADASIL) is the most frequent small artery brain disease caused by pathogenic variants of the NOTCH3 gene. During the disease, we still do not know how the various deficits progress and develop with each other at different stages of the disease. We aim to model disease progression, identify possible progressive subgroups and the effects of different covariates on clinical worsening. Methods: Data were from patients followed in the French CADASIL referral center, who were aged 25-80 years and had completed at least two visits and one of 14 clinical scores. Progression and variability were assessed using a Disease course model (Leaspy). A Gaussian mixture model was used to identify different progression subgroups. Logistic regressions were used to compare the characteristics between groups. Results: In 395 patients along 2007 visits, the follow-up ranged from 6 months to 19 years, with a mean of 7.5 years. They were 45% men with a mean age of 52.2 years. The evolution curves of the different scores showed that clinical manifestations develop heterogeneously and can vary considerably depending on the disease stage. We identified an early-onset, rapidly progressing subgroup of patients with earlier motor symptoms and focal neurological deficits, (median time-shift: 59 (Q1-Q3: 48.9-66.3), median acceleration rate: 0.84 (0.07-1.31), and a late-onset slowly progressing group, with earlier cognitive symptoms, (median time-shift: 69.2 (63.4-75.1), median acceleration rate: -0.18 (-0.48-0.14). Male gender, a lower education level, hypertension, and the NOTCH3 pathogenic variant location within EGFr 1-6 were found associated with this group difference. Discussion: Our results suggest a gradual and heterogeneous decline in different clinical and cognitive performances over the lifetime of CADASIL patients. Two progression profiles, one rapid and early and the other, more delayed and slower, are possible after the onset of symptoms. Although a major limitation of our study is that the clusters were assessed post-hoc which may induce some bias. Overall, male gender, a low level of education, the pathogenic variant location in EGFr 1 to 6 domains, smoking and/or arterial hypertension may affect the clinical progression of the disease.

1 Abstract

2

3 Background and Objectives: Cerebral Autosomal Dominant Arteriopathy with Subcortical  
4 Infarcts and Leukoencephalopathy (CADASIL) is the most frequent small artery brain disease  
5 caused by pathogenic variants of the NOTCH3 gene. During the disease, we still do not know  
6 how the various deficits progress and develop with each other at different stages of the  
7 disease. We aim to model disease progression, identify possible progressive subgroups and  
8 the effects of different covariates on clinical worsening.

9 Methods: Data were from patients followed in the French CADASIL referral center, who were  
10 aged 25-80 years and had completed at least two visits and one of 14 clinical scores.

11 Progression and variability were assessed using a Disease course model (Leaspy). A Gaussian  
12 mixture model was used to identify different progression subgroups. Logistic regressions  
13 were used to compare the characteristics between groups.

14 Results: In 395 patients along 2007 visits, the follow-up ranged from 6 months to 19 years,  
15 with a mean of 7.5 years. They were 45% men with a mean age of 52.2 years. The evolution  
16 curves of the different scores showed that clinical manifestations develop heterogeneously  
17 and can vary considerably depending on the disease stage. We identified an early-onset,  
18 rapidly progressing subgroup of patients with earlier motor symptoms and focal neurological  
19 deficits, (median time-shift: 59 (Q1-Q3: 48.9-66.3), median acceleration rate: 0.84 (0.07-  
20 1.31), and a late-onset slowly progressing group, with earlier cognitive symptoms, (median  
21 time-shift: 69.2 (63.4-75.1), median acceleration rate: -0.18 (-0.48-0.14). Male gender, a  
22 lower education level, hypertension, and the NOTCH3 pathogenic variant location within  
23 EGFr 1-6 were found associated with this group difference.

24 Discussion: Our results suggest a gradual and heterogeneous decline in different clinical and  
25 cognitive performances over the lifetime of CADASIL patients. Two progression profiles, one  
26 rapid and early and the other, more delayed and slower, are possible after the onset of  
27 symptoms. Although a major limitation of our study is that the clusters were assessed post-  
28 hoc which may induce some bias. Overall, male gender, a low level of education, the  
29 pathogenic variant location in EGFr 1 to 6 domains, smoking and/or arterial hypertension  
30 may affect the clinical progression of the disease.

31

## 32 Introduction

33

34 Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and  
35 Leukoencephalopathy (CADASIL) is the most frequently inherited cerebral small vessel  
36 disease (cSVD). It is caused by cysteine missense pathogenic variants in one of the 34  
37 epidermal growth-factor-like repeat (EGFr) domains of the NOTCH3 protein, a receptor  
38 located at the surface of smooth muscle cells in the wall of arterioles or at the surface of  
39 pericytes in capillaries.<sup>1</sup> Such pathogenic variants presumably lead to aggregation of both the  
40 mutant and wild forms of NOTCH3-ECD<sup>2</sup> with other components of the matrisome within the  
41 vascular wall<sup>3</sup>. The clinical spectrum of CADASIL is wide and includes attacks of migraine  
42 with aura, stroke, mood disturbances and diverse neuropsychiatric symptoms, cognitive  
43 impairment ranging from executive dysfunction up to severe dementia<sup>4</sup> and motor disability.  
44 All these clinical manifestations have already been reported with varying frequency, in many  
45 other cSVDs. Thus, CADASIL is now considered a unique model for better understanding the  
46 natural history and its underlying mechanisms in such conditions, that may span over  
47 multiple decades.

48 Accumulating data from cross-sectional or rare and short-term longitudinal studies  
49 allows to better delineate how the main manifestations of the disease develop over time. A  
50 reduction of processing speed is likely the earliest manifestation of cognitive decline during  
51 the disease course. Altered performances in executive functions such as attention,  
52 concentration or short-term and working memory may develop just after, also at an early  
53 stage. Later, modifications in all cognitive domains presumably increase gradually up to  
54 diffuse cognitive impairment with dementia at the latest stage.<sup>5,6,7</sup> Often, motor  
55 disturbances, with gait and balance problems develop slowly at first, then more obviously, in  
56 parallel with cognitive decline. These manifestations are often complicated by mood  
57 disturbances or neuropsychiatric symptoms, or by behavioral difficulties such as apathy.  
58 Previous studies have shown that these manifestations vary in intensity over time and in  
59 different groups of patients.<sup>8,9</sup> Over the last few years, we have also learned that at the same  
60 age, clinical severity can vary widely between individuals, depending particularly on certain  
61 vascular risk factors and the location of NOTCH3 pathogenic variants within the EGF domains.  
62 However, although our global vision of the disease has become more apparent, the exact  
63 natural history of the disease remains nonetheless imprecise. Thus, we don't know exactly  
64 how each clinical deficit develops in relation to the others, how they worsen over time, or  
65 how they combine during the different disease stages. This information is crucial, to  
66 determine the disease course variations from patient to patient and to ultimately enrich the  
67 information that will ease considerably therapeutic development by selecting the right  
68 patients, at the right time, using the appropriate assessment tools. In the present study, we  
69 aimed to determine how the different clinical manifestations of CADASIL are appearing and  
70 progressing over time and to what extent the disease course could differ between  
71 individuals. We hypothesized that an innovative analytical approach applied to a large cohort  
72 of patients could help answer these questions.

73

74 **Methods**

75

76 **Population selection**

77

78 Data were obtained from patients recruited at the French National Referral Centre  
79 CERVCO. In this study, all patients were included after confirmation of CADASIL diagnosis by a  
80 genetic test showing a cysteine pathogenic variant in the EGFR domains of the NOTCH3 gene.  
81 Participation in the cohort was systematically offered to all individuals who, at the request of  
82 their general practitioner or neurologist, were referred to the national CERVCO center in  
83 Paris for their management and follow-up. After obtaining their written consent, all patients  
84 were evaluated regularly by the same experts of the disease (DH, SG, NA and HC) and  
85 experienced neuropsychologists (AJ, SR, CM) using a unique case report form. They were  
86 consecutively enrolled with a follow-up systematically proposed to each participant  
87 approximately every two years, including a complete clinical evaluation, cognitive tests and  
88 MRI examination. The corresponding data were recorded using a secure web application to a  
89 common database (REDCAP). We selected patients from the whole cohort based on the  
90 following criteria: 1) age at inclusion between 25 and 80 years, 2) completion of at least two  
91 visits, 3) follow-up data including at least one clinical score among the 14 scores recorded  
92 during each visit. We included all patients who met these three criteria. The clinical  
93 information was collected by expert neurologists (NA, DH, SG and H.C and cognitive scores  
94 measured by experienced neuropsychologists (AJ, SR, CM)<sup>10,11</sup>.

95

96 **Clinical outcomes**

97

98 The clinical scores assess global cognitive efficiency, executive functions and memory  
99 performances, motor disability, focal neurological deficits, daily activities as well as the life  
100 quality. More specifically, cognitive efficiency was assessed using the Mattis Dementia Rating  
101 Scale (MDRS)<sup>12</sup> and the Vascular Dementia Assessment Scale cognitive scale (VADAS-Cog)<sup>13,14</sup>.  
102 The MDRS consists of 5 subscales related to attention, initiation/preservation, construction,  
103 conceptualization, and memory. Here only the initiation/perseveration subscore was  
104 considered for analysis. The VADAS-Cog is an extended version of the Alzheimer's Disease  
105 Assessment Scale-Cognitive subtest (ADAS-Cog) and includes additional subtests covering  
106 attention, working memory, executive function, and verbal fluency. Three of the VADAS-Cog  
107 components were included: the Digit Cancellation Test, the Symbol Digit Test, and the  
108 Backward Digit Span. We also obtained 3 scores from the Trail Making Test (TMT)<sup>15</sup>  
109 evaluating cognitive speed and mental flexibility, the TMT A time, TMT B time, and TMT B  
110 errors. Memory performance was assessed using 3 scores from the Free and Cued Selective  
111 Reminding Test adapted from the Grober and Buschke (GB) procedure:<sup>16</sup> the Total Free  
112 Recall, Index of Sensitivity to Cueing and Delayed Total Recall. The severity of disability was  
113 assessed using the modified Rankin scale and included the value of 6 (death) when the time  
114 at death was detailed. Focal neurological deficits were summarized using the NIH Stroke  
115 Score (NIHSS)<sup>17</sup> and limitations in daily activities by the Barthel index<sup>18</sup>. Finally, the quality of  
116 life was assessed using the Visual Analog Scale of the 3-level EuroQol (EQ VAS)<sup>19</sup>. In the  
117 present study scores were transformed and normalized in order to always increase from 0 to  
118 1 using their theoretical limits given in Table 2.

119

120 **Baseline covariates**

121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166

Based on a previous study<sup>9</sup>, various baseline parameters were considered as potential risk factors for more aggressive disease progression. They included: 1) gender (male or female), 2) education level classified as high (more than 13 years of education) or low otherwise, 3) alcohol consumption according to 3 categories: never, <2 glasses of wine per day for men (<1 glass or equivalent for women), and above these thresholds, 4) presence or absence of each of the following cardiovascular risk factors, hypercholesterolemia, diabetes, hypertension and smoking (yes for active /no for past or never), 5) presence of at least one cardiovascular risk factor (0 or  $\geq 1$ ), 6) number of cardiovascular risk factors (0, 1, 2, or >2), 6) the pathogenic variant's position located within EGF domains 1-6 of the NOTCH3 protein or within domains 7-34.<sup>20</sup>

### Mixed effects statistical model (Leaspy)

We used the Disease Course Mapping model<sup>21</sup>, a mixed-effects model implemented in an open-source Python library (*Leaspy standing for LEarning Spatiotemporal Patterns in Python*) which has been already tested for delineating multidimensional aspects of disease progression, in neurodegenerative diseases<sup>22,23,24</sup>. This innovative approach allows monitoring how the different clinical scores progress over time and in comparison to each other. It can also describe the disease trajectory at group and individual levels. Repeated clinical scores obtained longitudinally were considered for analysis.

The average disease progression was constructed as a mixed logistic curve model based on all score values normalized from 0 (best value) to 1 with subject level variability considered as the random effect. The average curve for each outcome was described by the parameters  $p_0$ ,  $v_0$  and  $t_0$ , where  $p_0$  and  $v_0$  are the position and velocity (derivative of the curve) at time  $t_0$ , the midpoint of the logistic. For multivariate analysis outcomes were modeled as a 14-point vector with age as the only regressor. Thus, from the mean trajectory, we could derive the individual trajectories based on 3 subject-specific parameters: 1) the time-shift  $\tau$  (tau) corresponding to the estimated time lag compared to the average progression (in years), indicating an earlier ( $\tau < \tau$  mean) or later disease onset ( $\tau > \tau$  mean), 2) the progression rate  $\xi$  (xi) indicating acceleration ( $\xi > 0$ ) or deceleration ( $\xi < 0$ ) compared to the average progression, 3) the spatial profile was defined by the inter-marker spacing parameters  $\omega$  (omega) accounting for the variable ordering within the sample. For each patient, one  $\omega$  value was related to each clinical score and indicated that a given score has begun to deteriorate before ( $\omega < 0$ ) or after ( $\omega > 0$ ) the average variation estimated in the whole population.

Finally, since there could be common patterns within a population, we used another parameter called "sources" for considering the degrees of freedom related to the sequence of events. As the number of sources was smaller than the number of omegas, this additional parameter simplified the estimation process. To interpret the results, the omegas were retrieved by an independent component analysis.

To evaluate the goodness-of-fit of the model to the actual data, we compared the estimated value of each visit with its observed value to compute the  $R^2$ .

### Definitions of subgroups of patients

167 To investigate which risk factors might influence the severity of the disease, we  
168 performed Mann-Whitney U tests on the individual parameters describing the patients'  
169 spatiotemporal profiles ( $\tau$ ,  $\xi$ , and  $\omega$ ) according to baseline covariates. Thereafter, we applied  
170 a Gaussian Mixture Model<sup>25</sup> to identify subgroups of patients having similar trajectories. For  
171 this purpose, we considered the individual parameters ( $\tau$ ,  $\xi$ , and the sources to keep the  
172 number of variables to a minimum) and defined the number of clusters according to the  
173 Akaike and the Bayesian Information Criteria. We repeated the algorithm 1000 times to  
174 stabilize the results. After the clustering process, to better describe the profile of patients  
175 belonging to different subgroups, we used Mann-Whitney U tests to compare the  
176 parameters  $\tau$ ,  $\xi$  and  $\omega$  between the subgroups. We also compared the RMSE (Root Mean  
177 Square Error) between the two subgroups as an additional performance metric of our  
178 disease course mapping model.

179 Finally, the effects of previous statistically significant baseline covariates were also  
180 assessed a posteriori, using a univariate logistic regression, followed by backward elimination  
181 in a multivariable model to best describe the different subgroups using as a dependent  
182 variable the subgroup label (early vs late). Independence between covariates was tested  
183 using  $\chi^2$  tests and interactions were added in the univariate models in case of association.  
184 Statistical tests were performed at the conventional two-tailed type I error of 0.05 using  
185 Python version 3.12.2.

186

#### 187 Standard Protocol Approvals, Registrations, and Patient Consents

188

189 Informed consent was obtained from each subject or from a close relative if  
190 necessary. The study was approved by an independent ethics committee (CEEI-IRB-17/388).

191

#### 192 Data Availability

193

194 Raw data may be shared (anonymized) at the request of any qualified investigator.

195 The script used for data analysis is available on github:

196 <https://github.com/KaisaridiSofia/CADASIL.git>

197



198 **Results**

199

200 **Population Description**

201

202 Clinical data were collected from 395 patients during 2007 visits, with a median of 4  
203 visits per patient (IQR: 3-7, Range: 2-14), with 66% having their NOTCH3 pathogenic variant  
204 located in EGFr high domains from 1 to 6 (Table 1).

205 The follow-up duration varied from 6 months to 19 years with a mean of 7.5 years  
206 (IQR: 3.5-11.1). The mean age of patients at inclusion was 52.2 years (IQR: 44.5-61). They  
207 were 45% male, 72% had a low educational background, 18% were current smokers, 21%  
208 hypertensive and 22.3% had no cardiovascular risk factor, while 23.7% had more than two  
209 (Supplementary Figure 1).

210

211 **CADASIL course mapping**

212

213 In the present study, we used 14 clinical scores collected longitudinally with 3 to 43%  
214 of missing data depending on the outcome (Supplementary Table 1). The best performing  
215 Disease Course model (Supplementary Figure 2) for delineating CADASIL progression showed  
216 that there were three groups of scores deteriorating at different stages of the disease. These  
217 groups were decided by a visual inspection over the population curves (Figure 1A) and were  
218 then confirmed using the normalized score value at the average inflection point of disease  
219 progression (Table 2). The first group of scores had normalized scores over 0.56, the second  
220 between 0.38 and 0.35 and the third group below 0.18. Cognitive scores were the first to  
221 deteriorate as different VADAS-Cog subscores largely exceeded 0.5 at the reference inflection  
222 point like the Symbol Digit Test, Backward Digit Span or Digit Cancellation Task (between  
223 0.76 and 0.87), while the TMT B Time and GB Total Free Recall were at 0.59 and 0.56  
224 respectively. The Backward Digit Span was however much higher than zero at the very start  
225 because only 5% of patients presented with the lowest corresponding score at time of  
226 inclusion. Conversely, the EQ-VAS quality of life (EQVAS), modified Rankin scale (mRS) and  
227 TMT A time (TMTA T) ranged from 0.35 to 0.38 at the reference inflection time, indicating  
228 that these clinical parameters were changing mainly after the intermediate stage of the  
229 disease. Finally, only later, the Mattis DRS Initiation scores, GB Index of Sensitivity to Cueing  
230 (GB Cueing), TMT B errors (TMTBE), GB Delayed Total Recall (GB Delayed), Barthel index  
231 (Barthel) and NIHSS index (NIHSS) changed most. Their scores were below 0.2 at the  
232 inflection time, indicating that, they had not yet changed significantly.

233 The time shift, as reflected by  $\tau$  values, was found significantly smaller (earlier onset)  
234 in patients with male gender, low education level, smoking, hypertension and who had a  
235 NOTCH3 pathogenic variant located in EGFR domains 1-6 (Figure 2, Supplementary Table 2).

236 The progression rate, as reflected by  $\xi$  values, was also found significantly higher  
237 (faster rate) in men compared to women (Figure 2, Supplementary Table 2).

238 Finally, regarding the symptoms' order, as reflected by  $\omega$  values (Figure 3,  
239 Supplementary Table 2), the education level was found to influence significantly all clinical  
240 measures, except the patient's life quality. While the Barthel, NIHSS, and Rankin scores were  
241 found to change slightly earlier, all the cognitive scores varied later in poorly educated  
242 patients. Gender was also found to influence the disease progression; memory performances  
243 (as assessed by different GB sub-scores) were found to modify earlier in women, whereas the

244 quality of life or the NIHSS scores was changing earlier in men. Hypertension was also  
245 associated with earlier changes in the Rankin and NIHSS scores, but with later modifications  
246 of the GB Total Free Recall and VADAS-Cog Digit Cancellation Task.

247 Concerning the  $R^2$  (Supplementary Figure 7) we observe values ranging from 0.68 to  
248 0.95 for almost all the scores except TMTB errors (0.40), Quality of life (0.26) and VADAS-Cog  
249 Backward Digit Span (0.47).

250

#### 251 Subgroups of patients with similar trajectories

252 The Gaussian Mixture Model based on  $\tau$ ,  $\xi$ , and  $\omega$  values, identified 2 subgroups of  
253 patients (Supplementary Figure 3) with different clinical trajectories of CADASIL  
254 corresponding to early or late disease progressors. The first subgroup, labeled “*early*”,  
255 included 37% of the patients with a median tau value of 59 (Q1-Q3: 48.9-66.3, Range: 29.5-  
256 91.7) and a median  $\xi$  value of 0.84 (Q1-Q3: 0.07-1.31, Range: -1.65-3.12) (Supplementary  
257 Figure 4, Supplementary Table 3) i.e. with an early and fast disease progression. The  
258 remaining 63% of the patients belonged to the second subgroup, labeled “*late*”, with a  
259 median tau value of 69.2 (Q1-Q3: 63.4-75.1, Range: 45.4-89.9) and a median  $\xi$  value of -0.18  
260 (Q1-Q3: -0.48-0.14, Range: -1.85-0.97) (Supplementary Figure 4, Supplementary Table 3),  
261 indicating a later and slower disease progression. In comparison to the average progression  
262 observed in the whole population, the corresponding curves (Figure 1B) showed that scores  
263 changes occurred earlier (dotted line shifted to the left) and progressed more rapidly (slopes  
264 of the curves are steeper) in the “*early*” subgroup of patients. On the other hand, in the late  
265 subgroup, the score changes occurred later (dotted line slightly shifted to the right) and  
266 more slowly (the slopes are smoother) (Figure 1C). The early subgroup had also smaller  $\omega$   
267 values for the Rankin Index (-0.06 vs 0.04, p-value<0.001) and the Barthel Index (medians -  
268 0.03 vs 0.02, p-value<0.001) suggesting earlier disability and dependency. We also observed  
269 earlier focal deficits (possibly related to stroke events) in this subgroup as reflected by the  
270 distribution of  $\omega$  values for the NIHSS (-0.01 vs 0.01, p-value<0.001) (Figure 4A,  
271 Supplementary Table 3). Conversely, an earlier onset of cognitive score changes as shown by  
272 differences of  $\omega$  values for VADAS cognitive subscores was detected in the “*late*” subgroup  
273 (Supplementary Table 3). In contrast, the distributions of  $\omega$  values for the EQVAS index, GB  
274 Delayed Total Recall and Index of Sensitivity to Cueing did not differ between these  
275 subgroups (Figure 4C, Supplementary Table 3). All  $\omega$  distributions are detailed in the  
276 Supplementary Figure 5. The sensitivity for the model is 81%, specificity 40%, PPV 69%, NPV  
277 57%, LR+ 1.35, and LR- 0.475 (Supplementary Figure 6) . The RMSE is similarly distributed  
278 between the two groups (Supplementary Figure 8), with a slightly higher tendency for the  
279 early subgroup. Overall, the mean values stay below 0.2 for all the scores.

280 Finally, we tested the effects of these five covariates on early subgroup membership.  
281 In a univariate analysis, we found that patients who were male (OR: 2.86, 95% CI: [1.87-  
282 4.39]), less educated (OR: 2.44 [1.48-4.02]), with hypertension (OR: 1.73 [1.06-2.83]),  
283 hypercholesterolemia (OR: 1.63 [1.07-2.48]), or diabetes (OR: 2.77 [1.05-7.32]) and with a  
284 NOTCH3 pathogenic variant located in EGFr domains 1-6 (OR: 1.60 [1.03-2.51]) were more  
285 likely to belong to the “*early*” group (Table 3). Although there were some significant  
286 associations between these variables (Supplementary Table 4) the addition of an interaction  
287 term in the univariate model was never significant. Following a backward elimination, our

288 multivariable logistic model showed that men (OR: 2.86 [1.84-4.44]), with lower education  
289 (OR: 2.66 [1.58-4.53]), hypertension (OR: 1.82 [1.07-3.10]) and NOTCH3 pathogenic variant  
290 located in EGFr domains 1-6 of the pathogenic variant (OR: 1.84 [1.14-3.00]) were at  
291 increased risk to have an “early” disease progression (Table 3).  
292

293 **Discussion**

294

295 The results of this study provide the first fine chronological mapping of the most  
296 characteristic clinical manifestations during CADASIL. This mapping was based on thousands  
297 of visits obtained using the same assessment tools. The results confirm that the permanent  
298 manifestations observed in symptomatic patients develop in a heterogeneous way and vary  
299 considerably depending on the disease stage.

300 The reduction of cognitive performances observed with the VADAS-Cog subscores,  
301 TMT B time and Total Free Recall assessed during the Grober and Buschke procedure,  
302 confirms that modifications of executive functions are the earliest detectable manifestations.  
303 These findings agree with the reduction of cognitive speed and flexibility reported in young  
304 CADASIL adults, long before the occurrence of strokes.<sup>26</sup> The early reduction of total free  
305 recall suggest that memory performances could be affected early although memory is usually  
306 considered long preserved during the course of cSVDs.<sup>27,28</sup> These results are consistent with  
307 recent studies demonstrating that while long term memory recall is damaged at late stage,  
308 other aspects of memory performances, particularly related to attention or working memory  
309 could be early affected during the course of CADASIL.<sup>29,30,31</sup> At the intermediate stage of the  
310 disease, disability is developing, parallel to significant cognitive slowing leading to prolonged  
311 TMT-A time with changes in life quality. Later, when motor disability is installed, cognitive  
312 decline deepens and alterations in cognitive initiation, perseverations, and memory deficit  
313 without improvement by cueing develop as reflected by changes in the MDRS initiation and  
314 various subscores of the Grober and Buschke procedure. Finally, the Barthel and NIHSS  
315 scores show obvious alterations only in patients at the most advanced disease stage, and  
316 associated with developing of difficulties in daily activities. We evaluated the CADASIL course  
317 map by calculating the  $R^2$ , where we obtained satisfying results for all but three scores :  
318 TMTB errors, Quality of life and VADAS-Cog Backward Digit Span. Nevertheless, as our goal  
319 here is to better describe the course of the disease rather than improve the accuracy of our  
320 prediction, we kept these scores because they provide important information for the  
321 evolution of the disease.

322 A key finding of our study is also that in a large population of symptomatic CADASIL  
323 patients followed over decades, two different profiles of clinical progression are detected. In  
324 one subgroup, individuals present an early onset and a rapid clinical deterioration, they also  
325 develop rapidly focal deficits, motor disability and dependency. In another subgroup of  
326 patients, the clinical progression appears late and slow, and cognitive symptoms are  
327 occurring earlier. These findings suggest that the phenotypic variability of CADASIL might be  
328 more complex than expected and does not only rely on the time of onset and rate of  
329 progression, but also on the order of the different clinical manifestations during the disease  
330 course. This complexity is further illustrated by the results of the different memory  
331 performance tests which are differently affected according to the disease stage. While a  
332 decrease of Total Free Recall is detected earlier in patients with a late clinical progression,  
333 both the Delayed Total Recall and Index of Sensitivity to Cueing follow a similar decline in  
334 patients with early or late clinical worsening. Comparing the RMSE between the subgroups  
335 we see similarly satisfying low means and the slightly higher values for the early subgroup,  
336 can be explained by the unbalanced proportions (37% vs 63%).

337 Our results also showed that the disease manifests differently in patients who are  
338 male, less educated, hypertensive or smokers or who have a pathogenic variant located in  
339 the EGFr domains 1-6 of the Notch3 receptor. Male gender has already been shown to be

340 associated with more severe clinical manifestations in CADASIL.<sup>32</sup> The pathogenic variant  
341 location was also previously found to be a strong predictor of disease severity<sup>33</sup> although it  
342 could not explain alone the different clinical profiles<sup>34</sup> even combined with gender.<sup>35</sup>  
343 Cardiovascular risk factors have been also suggested to modulate the disease severity,<sup>36</sup>  
344 particularly hypertension.<sup>29</sup> The diagnosis of CADASIL is based on the discovery of the  
345 pathogenic variant. The presence of vascular risk factors in no way can exclude this  
346 pathogenicity. There is increasing evidence that vascular risk factors could modulate the  
347 phenotype of the disease, for example by accelerating its course. This is partly what we  
348 observe and confirm in the results of this study. Interestingly, this study also reveals that  
349 education level might interfere with the type of disease progression. The exact reasons for  
350 this remain unclear. We cannot rule out however that this is not partly related to the large  
351 influence of education on cognitive performances, particularly on those related to executive  
352 functions.<sup>37</sup> Indeed, the education level could have magnified the reduction of performances  
353 in some tests and revealed a higher sensitivity in particularly well-educated individuals. This  
354 might explain some of our results showing earlier cognitive changes in association with late  
355 disease progression in highly educated individuals. These results also support that early  
356 cognitive changes does not mean necessarily, rapid clinical worsening with severe disability  
357 and dependency. Interestingly, our results also enabled us to estimate the time shift in  
358 relation to these covariates. The effects of gender, education level and smoking might  
359 correspond to a difference of 5 or 6 years. The effects of arterial hypertension or the location  
360 of the pathogenic variant would be responsible for a delay from 2 to 3 years. However, these  
361 results should be cautiously interpreted, given the likely bias in our cohort recruitment  
362 consulting a referral center with expert neurologists.

363 This study has multiple strengths. First, the amount of data, particularly the number  
364 of participants and measurements, but also the follow-up duration, were considerable, for  
365 such a rare disease. This was crucial to model the average population trajectory of cognitive  
366 and motor decline precisely. Second, our model has already been shown to be robust in the  
367 presence of missing data<sup>38</sup> which allowed us to consider all visits even when not all tests  
368 were available. Third, this model allowed us to examine how the scores were progressing  
369 with each other, which is rarely analyzed at a longitudinal level. Although the diversity of  
370 ages, onset times, and progression rates could make it difficult to develop a reliable model,  
371 our age reparameterization allowed us to map a population trajectory consistent with the  
372 disease course and to compare individuals who were at the same stage rather than at the  
373 same “biological” age. Fourth, a particularly strong asset was that we managed to study the  
374 disease progression at an individual level in addition to the population level. Fifth, the non-  
375 supervised approach revealed two groups of CADASIL patients having different trajectories  
376 not only in terms of temporal evolution but also in terms of the events sequence. The  
377 longitudinal profiling, which considered all clinical aspects of individual evolution, allowed us  
378 to detect patients who develop motor disability and dependency symptoms early, and  
379 patients developing cognitive symptoms first. Sixth, our results provide insight into how  
380 different covariates, known to influence the disease at baseline, could interfere with the  
381 disease progression.

382 There are also several limitations in this study. To integrate all the clinical information  
383 in the course of the disease we used scores that are coarse and have variable sensitivity. This  
384 heterogeneity might affect the accuracy of the model. Also, regarding the baseline

385 characteristics tested, unfortunately the database did not include enough information to be  
386 able to consider them longitudinally. Certainly, an external validation of the practical use of  
387 the clustering would add immense value to our results. For that a prospective validation  
388 study is needed and we hope that it can be carried out in the future. The individual clusters  
389 were obtained post-hoc using the individual parameters as estimated by our model. A  
390 clustering procedure integrated since the first estimation steps of the model and considering  
391 the scores changes at each stage might have provided a more insightful classification. At the  
392 population level, the present study focused on the point at which clinical scores began to  
393 change, whereas we could also have examined the different rates of change on the variation  
394 curves of the clinical features throughout the trajectory of the disease. Moreover, the lack of  
395 external validation and the potential bias in the cohort recruitment pose additional  
396 challenges. Incorporating new observations, more importantly additional visits for some  
397 patients would help improve the model's performance. Furthermore, because all the  
398 available data was used in the training of the model, these represent best-case scenario  
399 estimates, the performance on new patients will likely be lower. A study aiming at prediction,  
400 using a richer dataset, split into training and test set could also result to more insightful  
401 conclusions. Finally, in this study, only deficits that persisted and worsened over time were  
402 assessed. The occurrence of stroke and its influence on clinical and cognitive worsening was  
403 not specifically studied and would require further evaluation. Future research with the  
404 potential use of imaging data could also help to improve detailed chronological mapping of  
405 disease progression.

406 In conclusion, using a disease progression model and a large amount of longitudinal  
407 data, we collected reliable and key clinical information concerning the course of CADASIL.  
408 Our results support that there are two groups of patients, slow and fast progressors, and that  
409 different factors could explain this discrepancy. They also help delineate both the dynamics  
410 and order of multiple clinical score changes during the disease progression which is crucial  
411 for preparing future clinical trials.

412

413 Acknowledgements:

414 This research was supported by the state aid managed by the Agence Nationale de la  
415 Recherche under the France 2030 program (grant RHU TRT\_cSVD, ANR:16-RHUS-0004), by  
416 the Association ARNEVA (Association de Recherche en Neurologie Vasculaire) and by the  
417 European CADANHIS funding (EJPRD 2023).

418

419 References

- 420 1. Rutten JW, Hack RJ, Duering M, et al. Broad phenotype of cysteine-altering NOTCH3  
421 variants in UK Biobank: CADASIL to nonpenetrance. *Neurology*. 2020;95(13):e1835–  
422 e1843. doi:[10.1212/WNL.0000000000010525](https://doi.org/10.1212/WNL.0000000000010525)
- 423 2. Dupre N, Gueniot F, Domenga-Denier V, et al. Protein aggregates containing wild-type  
424 and mutant NOTCH3 are major drivers of arterial pathology in CADASIL. *J Clin Invest*.  
425 2024;134(8):e175789. doi:[10.1172/JCI175789](https://doi.org/10.1172/JCI175789)
- 426 3. Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular  
427 matrisome: A convergent mechanism in small vessel disease of the brain?. *J Cereb*  
428 *Blood Flow Metab*. 2016;36(1):143-157. doi:[10.1038/jcbfm.2015.62](https://doi.org/10.1038/jcbfm.2015.62)
- 429 4. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserre E, Bousser MG. Cadasil. *Lancet*  
430 *Neurol*. 2009;8(7):643–653. doi: [https://doi.org/10.1016/s1474-4422\(09\)70127-9](https://doi.org/10.1016/s1474-4422(09)70127-9)
- 431 5. Amberla K, Wäljas M, Tuominen S, et al. Insidious cognitive decline in CADASIL.  
432 *Stroke*. 2004;35(7):1598-1602. doi:[10.1161/01.STR.0000129787.92085.0a](https://doi.org/10.1161/01.STR.0000129787.92085.0a)
- 433 6. Peters N, Opherck C, Danek A, Ballard C, Herzog J, Dichgans M. The pattern of cognitive  
434 performance in CADASIL: a monogenic condition leading to subcortical ischemic  
435 vascular dementia. *Am J Psychiatry*. 2005;162(11):2078-2085.  
436 doi:[10.1176/appi.ajp.162.11.2078](https://doi.org/10.1176/appi.ajp.162.11.2078)
- 437 7. Brookes RL, Hollocks MJ, Tan RY, Morris RG, Markus HS. Brief Screening of Vascular  
438 Cognitive Impairment in Patients With Cerebral Autosomal-Dominant Arteriopathy  
439 With Subcortical Infarcts and Leukoencephalopathy Without Dementia. *Stroke*.  
440 2016;47(10):2482-2487. doi:[10.1161/STROKEAHA.116.013761](https://doi.org/10.1161/STROKEAHA.116.013761)
- 441 8. Brice S, Jabouley A, Reyes S, et al. Modeling the cognitive trajectory in CADASIL. *J*  
442 *Alzheimers Dis*. 2020;77(1):291–300. doi:[10.3233/JAD-200310](https://doi.org/10.3233/JAD-200310)
- 443 9. Brice S, Reyes S, Jabouley A, et al. Trajectory Pattern of Cognitive Decline in Cerebral  
444 Autosomal Dominant Arteriopathy With Subcortical Infarcts and  
445 Leukoencephalopathy. *Neurology*. 2022;99(10):e1019–e1031.  
446 doi:[10.1212/WNL.0000000000200805](https://doi.org/10.1212/WNL.0000000000200805)
- 447 10. Viswanathan A, Guichard JP, Gschwendtner A, et al. Blood pressure and haemoglobin  
448 A1c are associated with microhaemorrhage in CADASIL: a two-centre cohort study.  
449 *Brain*. 2006;129(Pt9):2375-2383. doi:[10.1093/brain/awl177](https://doi.org/10.1093/brain/awl177)
- 450 11. Chabriat H, Hervé D, Duering M, et al. Predictors of Clinical Worsening in Cerebral  
451 Autosomal Dominant Arteriopathy With Subcortical Infarcts and  
452 Leukoencephalopathy: Prospective Cohort Study. *Stroke*. 2016;47(1):4-11.  
453 doi:[10.1161/STROKEAHA.115.010696](https://doi.org/10.1161/STROKEAHA.115.010696)
- 454 12. Shay KA, Duke LW, Conboy T, Harrell LE, Callaway R, Folks DG. The clinical validity of  
455 the Mattis Dementia Rating Scale in staging Alzheimer’s dementia. *J*  
456 *Geriatr Psychiatry Neurol*. 1991;4(1):18-25. doi:[10.1177/089198879100400104](https://doi.org/10.1177/089198879100400104)

- 457 13. Madureira S, Verdelho A, Ferro J, et al. Development of a neuropsychological battery  
458 for the Leukoaraiosis and Disability in the Elderly Study (LADIS): experience and  
459 baseline data. *Neuroepidemiology*. 2006;27(2):101-116. doi:[10.1159/000095381](https://doi.org/10.1159/000095381)  
460 14. Ferris SH. General measures of cognition. *Int Psychogeriatr*. 2003;15 Suppl 1:215-217.  
461 doi:[10.1017/S1041610203009220](https://doi.org/10.1017/S1041610203009220)  
462 15. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. *Nat*  
463 *protoc*. 2006;1(5):2277–2281. doi:[10.1038/nprot.2006.390](https://doi.org/10.1038/nprot.2006.390)  
464 16. Epelbaum S, Benisty S, Reyes S, et al. Verbal memory impairment in subcortical  
465 ischemic vascular disease: a descriptive analysis in CADASIL. *Neurobiol Aging*.  
466 2011;3(12):2172-2182. doi:[10.1016/j.neurobiolaging.2009.12.018](https://doi.org/10.1016/j.neurobiolaging.2009.12.018)  
467 17. Adams HPJ, Davis PH, Leira EC, et al. Baseline NIH Stroke Scale score strongly predicts  
468 outcome after stroke: A report of the Trial of Org 10172 in Acute Stroke Treatment  
469 (TOAST). *Neurology*. 1999;53(1):126-131. doi:[10.1212/wnl.53.1.126](https://doi.org/10.1212/wnl.53.1.126)  
470 18. Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE. Stroke rehabilitation:  
471 analysis of repeated Barthel index measures. *Arch Phys Med Rehabil*. 1979;60(1):14–  
472 17.  
473 19. Naess H, Lunde L, Brogger J. The effects of fatigue, pain, and depression on quality of  
474 life in ischemic stroke patients: the Bergen Stroke Study. *Vasc Health Risk Manag*.  
475 2012;8:407-413. doi: [10.2147/VHRM.S32780](https://doi.org/10.2147/VHRM.S32780)  
476 20. Rutten JW, Van Eijsden BJ, Duering M, et al. The effect of NOTCH3 pathogenic variant  
477 position on CADASIL disease severity: NOTCH3 EGFr 1-6 pathogenic variant are  
478 associated with a more severe phenotype and lower survival compared with EGFr 7-  
479 34 pathogenic variant [published correction appears in *Genet Med*. 2018 Sep 20;:].  
480 *Genet Med*. 2019;21(3):676-682. doi:[10.1038/s41436-018-0088-3](https://doi.org/10.1038/s41436-018-0088-3)  
481 21. Schirratti JB, Allasonnière S, Colliot O, Durrleman S. A Bayesian Mixed- Effects Model  
482 to Learn Trajectories of Changes from Repeated Manifold-Valued Observations. *J*  
483 *Mach Learn Res*. 2017;18:1-33  
484 22. Ortholand J, Pradat PF, Tezenas du Montcel S, Durrleman S. Interaction of sex and  
485 onset site on the disease trajectory of amyotrophic lateral sclerosis. *J Neurol*.  
486 2023;270(12):5903-5912. doi:[10.1007/s00415-023-11932-7](https://doi.org/10.1007/s00415-023-11932-7)  
487 23. Koval I, Bône A, Louis M, et al. AD Course Map charts Alzheimer's disease  
488 progression. *Sci Rep*. 2021;11(1):8020. Published 2021 Apr 13. doi:[10.1038/s41598-021-87434-1](https://doi.org/10.1038/s41598-021-87434-1)  
489 24. Koval I, Dighiero-Brecht T, Tobin AJ, et al. Forecasting individual progression  
490 trajectories in Huntington disease enables more powered clinical trials. *Sci Rep*.  
491 2022;12(1):18928. Published 2022 Nov 7. doi:[10.1038/s41598-022-18848-8](https://doi.org/10.1038/s41598-022-18848-8)  
492 25. Zhou Z-H. *Machine Learning*. Springer Singapore. 2021:222-227  
493 26. Benisty S, Reyes S, Godin O, et al. White-matter lesions without lacunar infarcts in  
494 CADASIL. *J Alzheimers Dis*. 2012;29(4):903-911. doi: [10.3233/JAD-2012-111784](https://doi.org/10.3233/JAD-2012-111784)  
495 27. Kalimo H, Ruchoux MM, Viitanen M, Kalaria RN. CADASIL: a common form of  
496 hereditary arteriopathy causing brain infarcts and dementia. *Brain Pathol*.  
497 2002;12(3):371-384. doi:[10.1111/j.1750-3639.2002.tb00451.x](https://doi.org/10.1111/j.1750-3639.2002.tb00451.x)  
498 28. Dichgans M. Cognition in CADASIL. *Stroke*. 2009;40(3 Suppl):S45-S47.  
499 doi:[10.1161/STROKEAHA.108.534412](https://doi.org/10.1161/STROKEAHA.108.534412)  
500 29. Ciolli L, Pescini F, Salvadori E, et al. Influence of vascular risk factors and  
501 neuropsychological profile on functional performances in CADASIL: results from the  
502



- 503 Microvascular LEukoencephalopathy Study (MILES). *Eur J Neurol.* 2014;21(1):65-71.  
504 doi:[10.1111/ene.12241](https://doi.org/10.1111/ene.12241)
- 505 30. Jolly AA, Nannoni S, Edwards H, Morris RG, Markus HS. Prevalence and Predictors of  
506 Vascular Cognitive Impairment in Patients With CADASIL. *Neurology.*  
507 2022;99(5):e453-e461. Published 2022 Aug 1. doi:[10.1212/WNL.0000000000200607](https://doi.org/10.1212/WNL.0000000000200607)
- 508 31. Rendon J, Zuluaga Y, Velilla L, et al. Event-related potential correlates of recognition  
509 memory in asymptomatic individuals with CADASIL. *Brain Res.* 2019;1707:74-78.  
510 doi:[10.1016/j.brainres.2018.11.016](https://doi.org/10.1016/j.brainres.2018.11.016)
- 511 32. Gunda B, Hervé D, Godin O, et al. Effects of gender on the phenotype of CADASIL.  
512 *Stroke.* 2012;43(1):137-141. doi:[10.1161/STROKEAHA.111.631028](https://doi.org/10.1161/STROKEAHA.111.631028)
- 513 33. Dupé C, Guey S, Biard L, et al. Phenotypic variability in 446 CADASIL patients: Impact  
514 of NOTCH3 gene mutation location in addition to the effects of age, sex and vascular  
515 risk factors. *J Cereb Blood Flow Metab.* 2023;43(1):153-166.  
516 doi:[10.1177/0271678X221126280](https://doi.org/10.1177/0271678X221126280)
- 517 34. Zhang R, Ouin E, Grosset L, et al. Elderly CADASIL patients with intact neurological  
518 status. *J Stroke.* 2022;24(3):352-362. doi:[10.5853/jos.2022.01578](https://doi.org/10.5853/jos.2022.01578)
- 519 35. Lebenberg J, Guichard JP, Guillonnet A, et al. The Epidermal Growth Factor Domain of  
520 the Mutation Does Not Appear to Influence Disease Progression in CADASIL When  
521 Brain Volume and Sex Are Taken into Account. *AJNR Am J Neuroradiol.*  
522 2022;43(5):715-720. doi:[10.3174/ajnr.A7499](https://doi.org/10.3174/ajnr.A7499)
- 523 36. Adib-Samii P, Brice G, Martin RJ, Markus HS. Clinical spectrum of CADASIL and the  
524 effect of cardiovascular risk factors on phenotype: study in 200 consecutively  
525 recruited individuals. *Stroke.* 2010;41(4):630-634.  
526 doi:[10.1161/STROKEAHA.109.568402](https://doi.org/10.1161/STROKEAHA.109.568402)
- 527 37. Ekström I, Josefsson M, Bäckman L, Laukka EJ. Predictors of cognitive aging profiles  
528 over 15 years: A longitudinal population-based study. *Psychol Aging.* Published online  
529 May 16, 2024. doi: [10.1037/pag0000807](https://doi.org/10.1037/pag0000807)
- 530 38. Couronné R, Vidailhet M, Corvol J-C, Lehericy S, Durrleman S. Learning disease  
531 progression models with longitudinal data and missing values. *IEEE 16th International  
532 Symposium on Biomedical Imaging (ISBI).* 2019;1033-1037. doi:  
533 [10.1109/ISBI.2019.8759198](https://doi.org/10.1109/ISBI.2019.8759198)
- 534

535 **Tables**

536

537 **Table 1:** Baseline characteristics

Baseline characteristic	
Age	52.9 (44.5-610) [24.6-79.4]
Gender: male	178 (45%)
Education: <13 years	284 (72%)
Alcohol consumption	
never	119 (37%)
<2 glasses of wine for a man	166 (51%)
>2 glasses of wine for a man	38 (12%)
Smoking	71 (18%)
Hypertension	84 (21%)
Hypercholesteremia	150 (40%)
Diabetes	17 (4%)
Presence of at least one cardiovascular risk factor	307 (78%)
Number of cardiovascular risk factors	
0	88 (24%)
1	164 (44%)
2	93 (25%)
>2	27 (7%)
Pathogenic variants position : within EGF domains 1-6	259 (66%)
VADAS – Cog Symbol Digit Test	4 (1-8) [1-10]
VADAS – Cog Backward Digit Span	4 (2-5) [1-5]
VADAS – Cog Digit Cancellation Task	5 (2-8) [1-10]
TMT B Time	85 (60-137) [27-300]
GB Total Free Recall	29 (23-34.50) [0-47]
Modified Rankin Scale	1 (0-2) [0-5]
TMT A Time	37 (27-53.50) [13-180]
EQVAS Quality of life	75 (60-90) [0-100]
MDRS initiation	37 (32-37) [6-37]
GB Index of Sensitivity to Cueing	94.44 (84.21-100) [16.67-100]
GB Delayed Total Recall	16 (15-16) [1-16]
TMT B Errors	0 (0-1) [0-8]
NIHSS Index	0 (0-1) [0-25]
Barthel Index	100 (100-100) [0-100]

538 **Legend:** Median, interquartile range (in brackets) and ranges (in closed brackets ) are  
539 provided for quantitative variables. Frequencies and percentages (in brackets) are given for  
540 qualitative variables.

541

542  
543  
544

**Table 2:** Normalized score values obtained for each test at the average mid-time (tau mean) of disease progression

<i>Score (label) [evolution (best to worst)]</i>	<i>Normalized value at the average mid-time of disease progression (tau mean)</i>
VADAS-Cog Symbol Digit Test (Symbol-Digit) [1-10]	0.87
VADAS-Cog Backward Digit Span (Backward-Digit) [1-5]	0.82
VADAS-Cog Digit Cancellation Task (Digit-Cancel) [1-10]	0.76
Trail Making Test B Time (TMT-B T) [0-300]	0.59
GB Total Free Recall (GB-free) [48-0]	0.56
EQ VAS Quality of life (EQVAS) [100-0]	0.38
Modified Rankin Scale (Rankin) [0-6]	0.37
Trail Making Test A time (TMT-A T) [0-180]	0.35
MDRS Initiation (MDRS-initiation) [37-0]	0.18
GB Index of Sensitivity to Cueing (GB-cueing) [100 – 1]	0.12
Trail Making Test B errors (TMT-B E) [0-24]	0.05
GB Delayed Total Recall (GB-delayed) [16-0]	0.04
Barthel Index (Barthel) [100-0]	0.03
NIHSS index (NIHSS) [0-42]	0.03

545 **Legend:** The different tests employed in the study are detailed with their labels in brackets  
546 (subsequently used in Figure 1) and their limit values after their normalization. The second  
547 column indicates the normalized values of each score at the average time of inflection of  
548 disease progression obtained using the whole population (Figure 1A).

549 **Table 3: Risk factors of belonging to the earlier subgroup of disease progression**

550

<i>Baseline covariate</i>	<i>Univariate models</i>		<i>Multivariable model</i>	
	<i>Odds ratio (95% CI)</i>	<i>p-value</i>	<i>Odds ratio (95% CI)</i>	<i>p-value</i>
<b>Male gender</b>	<b>2.86 (1.87-4.39)</b>	<b>&lt;0.001</b>	<b>2.86 (1.84-4.44)</b>	<b>&lt;0.001</b>
<b>Education level</b>	<b>2.44 (1.48-4.02)</b>	<b>&lt;0.001</b>	<b>2.66 (1.58-4.53)</b>	<b>&lt;0.001</b>
Smoking	1.21 (0.71-2.03)	0.485		
<b>Hypertension</b>	<b>1.73 (1.06-2.83)</b>	<b>0.027</b>	<b>1.82 (1.07-3.10)</b>	<b>0.026</b>
<b>Hypercholesterolemia</b>	<b>1.63 (1.07-2.48)</b>	<b>0.021</b>		
<b>Diabetes</b>	<b>2.77 (1.05-7.32)</b>	<b>0.039</b>		
<b>Pathogenic variant location (EGFR 1-6)</b>	<b>1.60 (1.03-2.51)</b>	<b>0.036</b>	<b>1.84 (1.14-3.00)</b>	<b>0.013</b>
Low education level *Hypercholesterolemia	0.91 (0.31-2.61)	0.86		
Low education level *Pathogenic variant location (EGFR 1-6)	1.11 (0.33-3.63)	0.87		
Hypertension*Hypercholesterolemia	0.84 (0.31-2.27)	0.74		
Hypertension*Diabetes	0.82 (0.11-6.05)	0.84		
Hypertension*Pathogenic variant location (EGFR 1-6)	1.11 (0.40-1.13)	0.85		
Hypercholesterolemia*Diabetes	0.18 (0.02-2.12)	0.17		

551 **Legend:** The results of univariate logistic regression are shown with parameters selected  
552 from baseline covariates by Mann-Whitney U tests showing a significant difference in  
553 individual  $\eta$  or  $\xi$  values or a significant interaction by chi-2 tests (selection with  $p < 0.05$ , left).  
554 The results of the final multivariable model obtained using backward removal on p-values  
555 and the lowest AIC are also presented (right). The odds ratio with the corresponding 95%  
556 confidence interval and p-value are given for each variable. The significant results are in bold.  
557

558 **Figures**

559

560 **Figure 1**

561

562 **Title: CADASIL course map as delineated by the average longitudinal progression of 14**  
563 **clinical scores – 1A: Average disease progression in the whole population – 1B: Average**  
564 **disease progression in patients with early score changes – 1C: Average disease progression**  
565 **in patients with late score changes.**

566

567 **Legend:** Each colored curve represents the average progression of a specific clinical score  
568 (detailed on the right side) which was transformed and normalized for obtaining the same  
569 scale of progression from the best (0) to worst (1) score values. The effect of the time-shift  
570 and acceleration is to slide the points representing the patients' measurements along the  
571 curve so that the actual age of the subject at each visit moves to the corresponding disease  
572 age, represented in the x-axis. The vertical dotted lines ( $\tau$  mean) indicate the age at midpoint  
573 inflection of the average trajectory of patients score values. After obtaining the results of  
574 clustering analysis, the average progression was plotted for patients belonging to the group  
575 with an early clinical progression (1-B) and for those belonging to the group with the late  
576 onset respectively (1C).

577

578 **Figure 2**

579

580 **Title: Significant differences identified for temporal individual parameters  $\tau$  and  $\xi$  values**  
581 **according to baseline cohort parameters using Mann-Whitney tests**

582

583 **Legend:** Boxplots showing the distributions of individual time shift ( $\tau$ ) and progression rate  
584 ( $\xi$ ) of the disease in reference to the average disease progression curve according to the  
585 gender, education level, smoking, hypertension, and the pathogenic variant location in EGFR  
586 domains 1-6 versus 7-34. A greater  $\tau$  value in one group indicates a significant time shift in  
587 the disease progression i-e a later change of clinical scores in comparison to the other group.  
588 A greater  $\xi$  value or progression rate in one group indicates a faster disease progression  
589 compared to that observed in the other group. Significant results obtained with the Mann-  
590 Whitney U tests are indicated with a \*.

591

592 **Figure 3**

593

594 **Title: Significant differences observed in  $\omega$  values (order of clinical score changes)**  
595 **according to baseline cohort parameters using Mann-Whitney tests**

596

597 **Legend:** Boxplots showing the distributions of individual values of the inter-score spacing  $\omega$   
598 according to the education level, gender and prior history of hypertension. The x-axis  
599 corresponds to  $\omega$  values, and the 14 clinical scores are presented in the y axis. A smaller  $\omega$   
600 value for a given score in one group indicates that this score deteriorates earlier than for the  
601 other group and vice-versa. Only distributions with a significant group difference are shown  
602 (Mann-Whitney U tests), they were observed according to education level, gender and  
603 hypertension. No significant difference was detected in relation to smoking or to the NOTCH3  
604 Significant results obtained with the Mann-Whitney U tests are indicated with a \*.

605 **Figure 4**

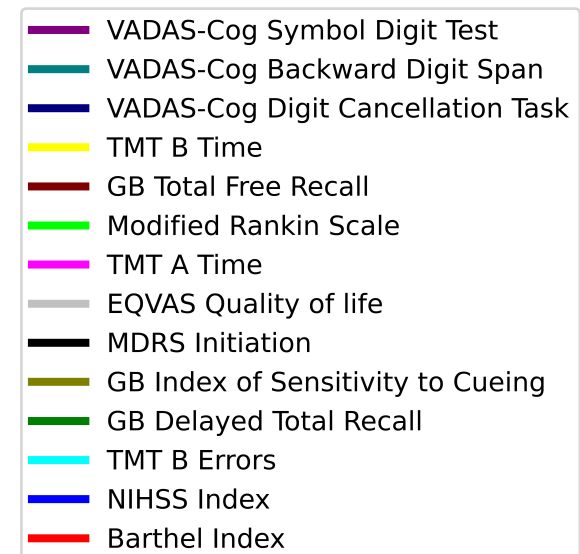
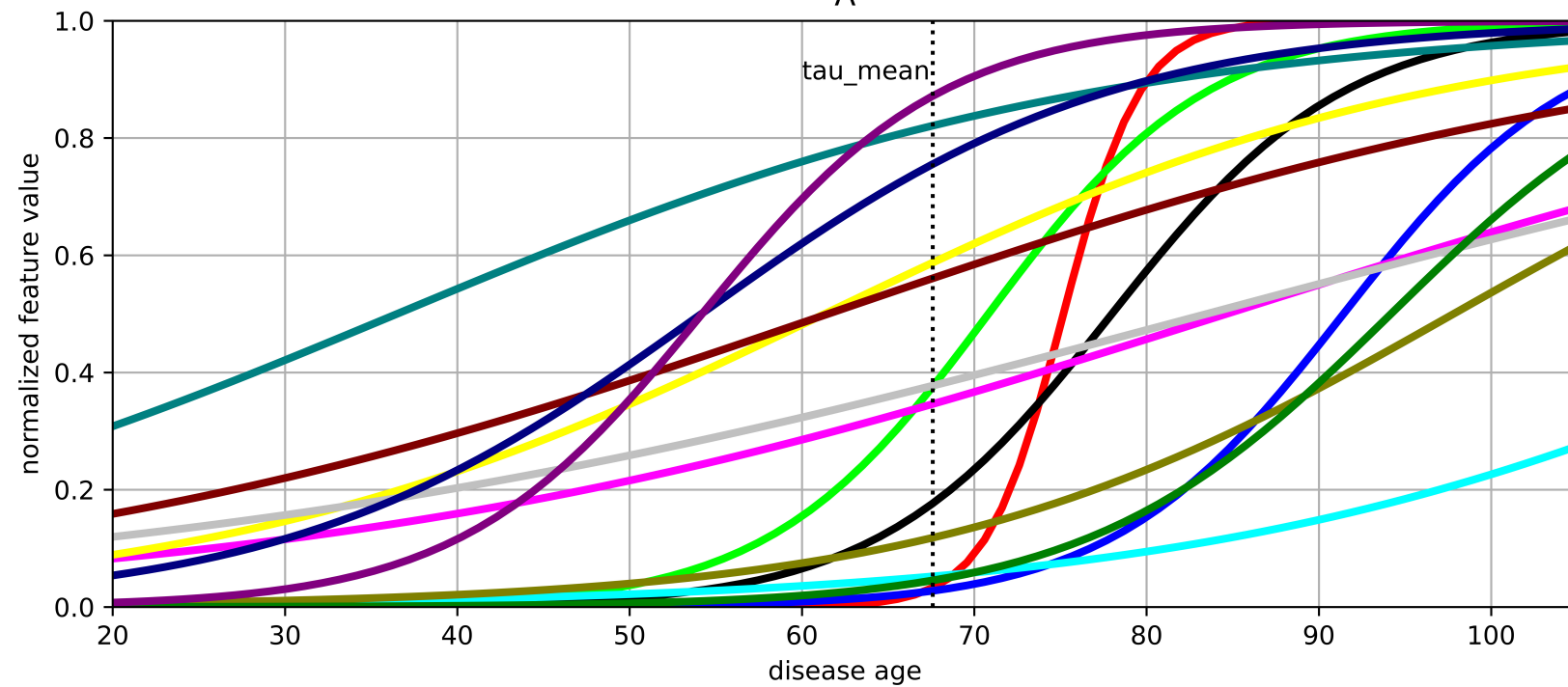
606

607 **Title: Distribution of individual  $\omega$  values for different clinical scores in the group with early**  
608 **and in the group with late disease progression**

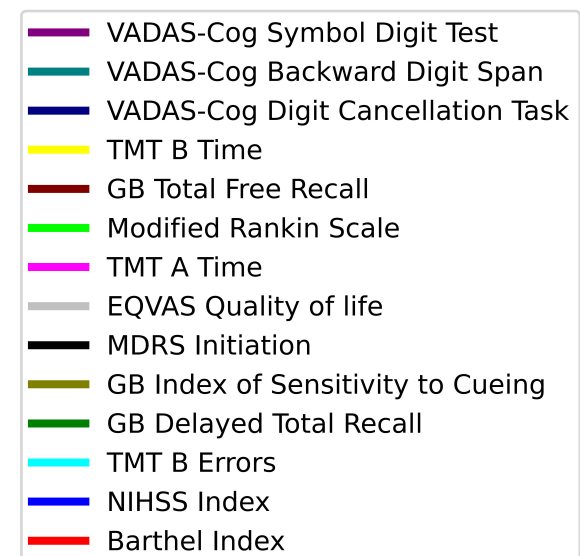
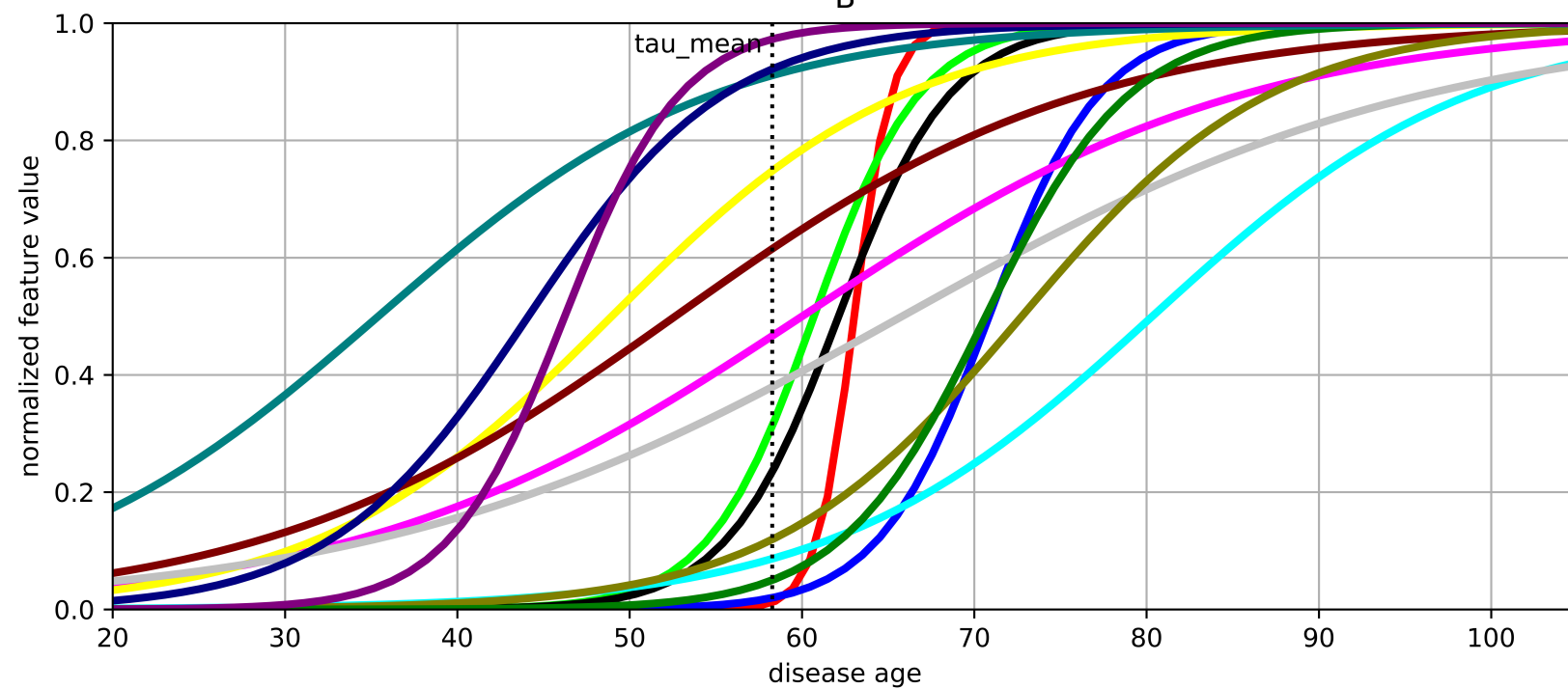
609

610 **Legend:** Each point corresponds to one patient. The x, y, and z axis correspond to the  
611 variation of  $\omega$  values for different clinical scores. A smaller  $\omega$  value detected in one group  
612 suggests that the corresponding clinical score deteriorates earlier than the other scores in  
613 comparison to what is observed in the other group and vice-versa. Figure A shows that  
614 scores related to motor disability, focal deficits and dependency (Rankin, NIH scores, Barthel  
615 index) deteriorate earlier than the other clinical changes in patients with early progression  
616 (red) compared to patients with late progression (green). Conversely, Figure B shows that the  
617 VADAS cognitive sub-scores modifications occur earlier than the other clinical changes in  
618 patients with late disease progression (green) compared to patients with early disease  
619 progression (red). In Figure C, both two memory scores derived from the Grober and Buske  
620 procedure (Index of Sensitivity to Cueing and Delayed Total Recall) as well as the quality of  
621 life index seem to develop at the same time then the other clinical manifestations in the two  
622 groups of early or late disease progression.

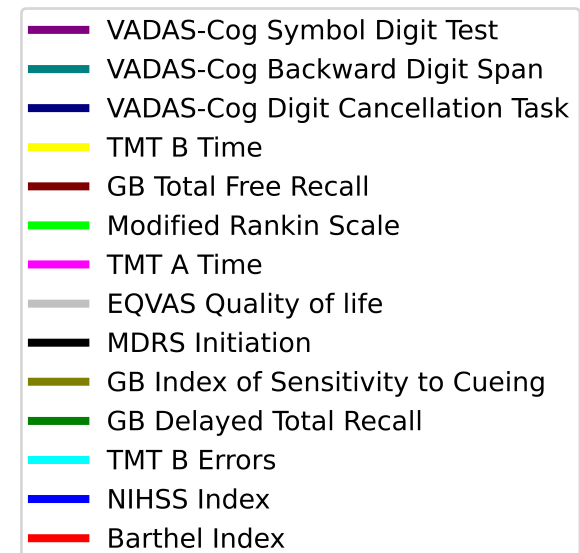
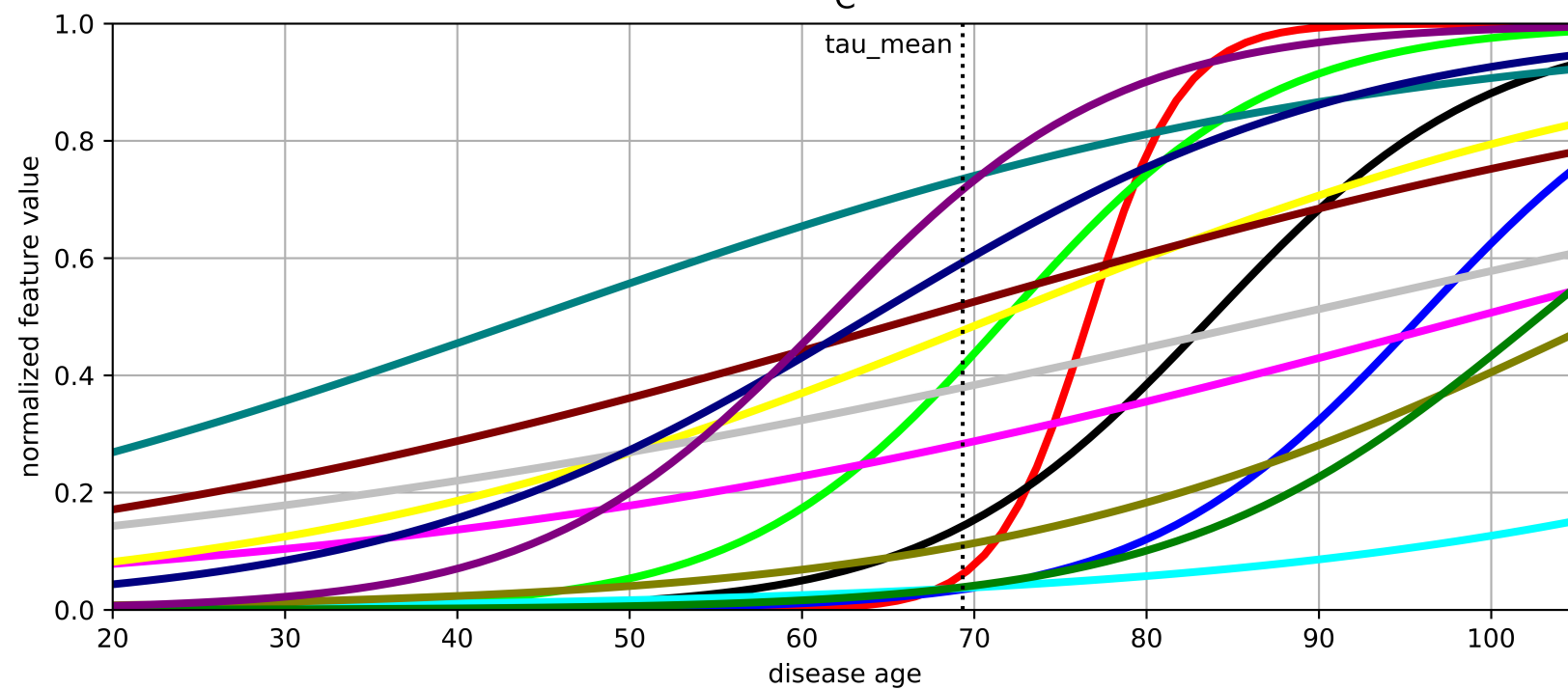
A

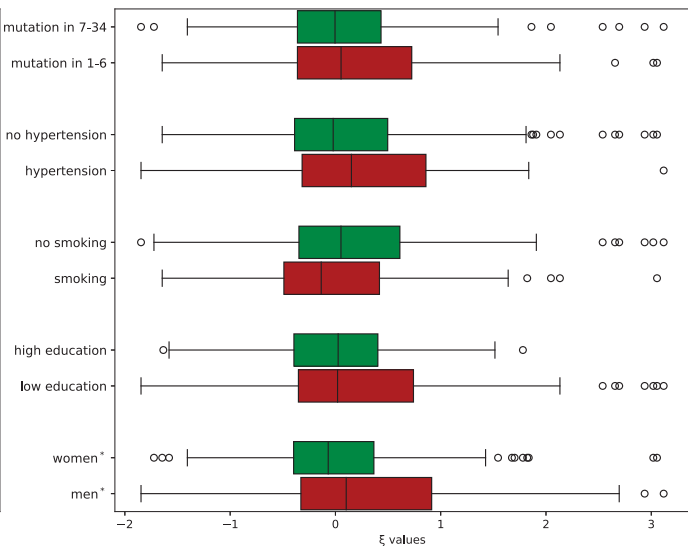
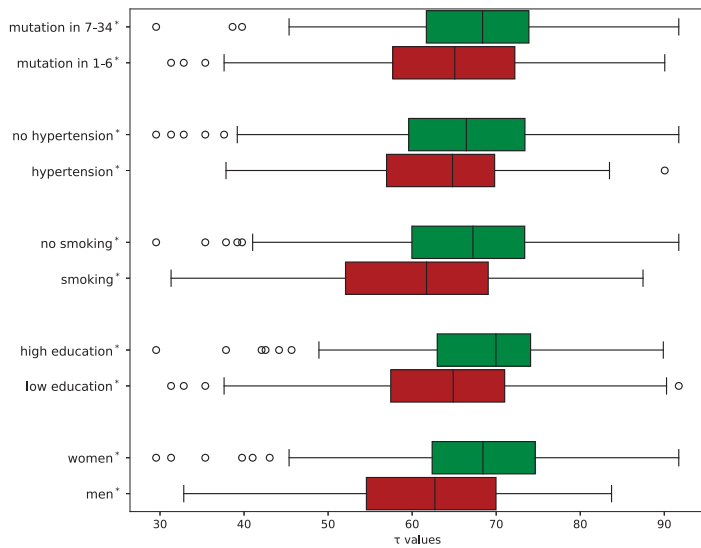


B

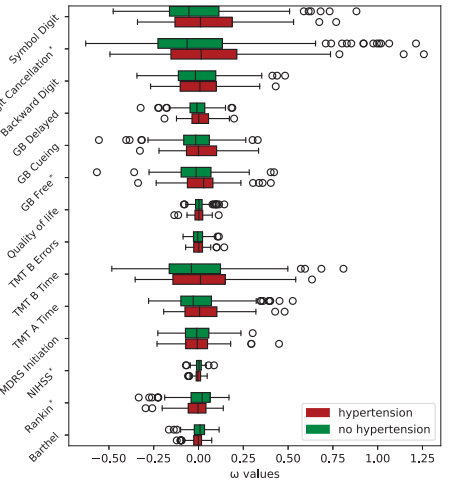
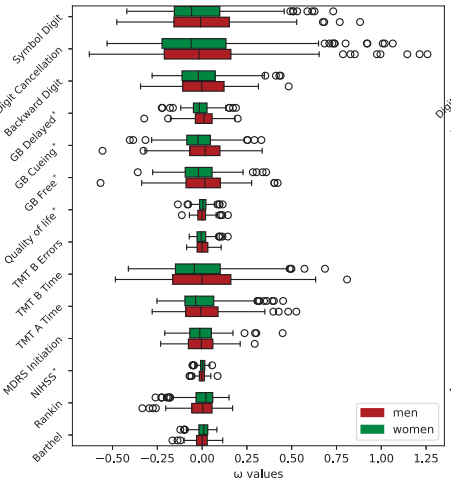
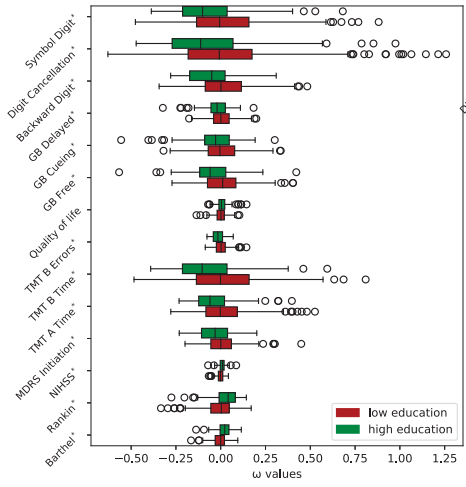


C



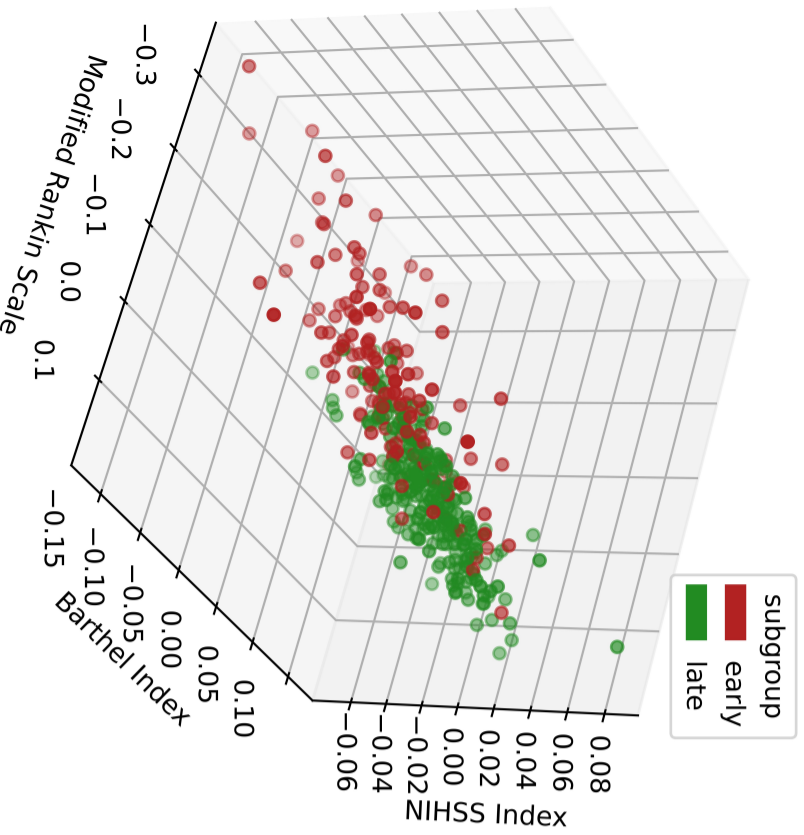




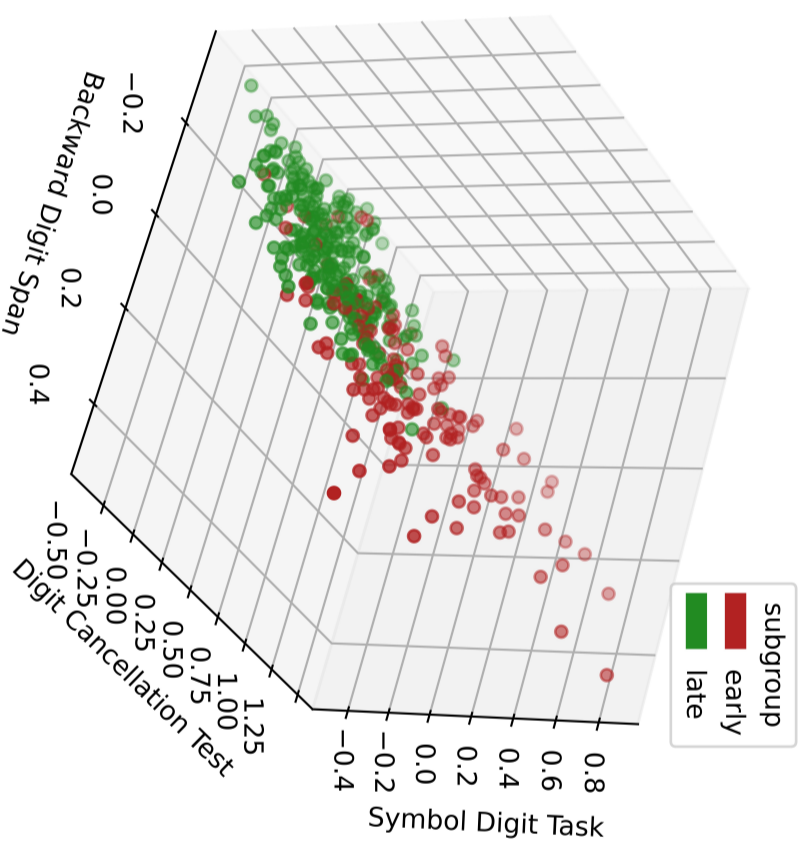


# Intermarker spacing parameters - $\omega$ values

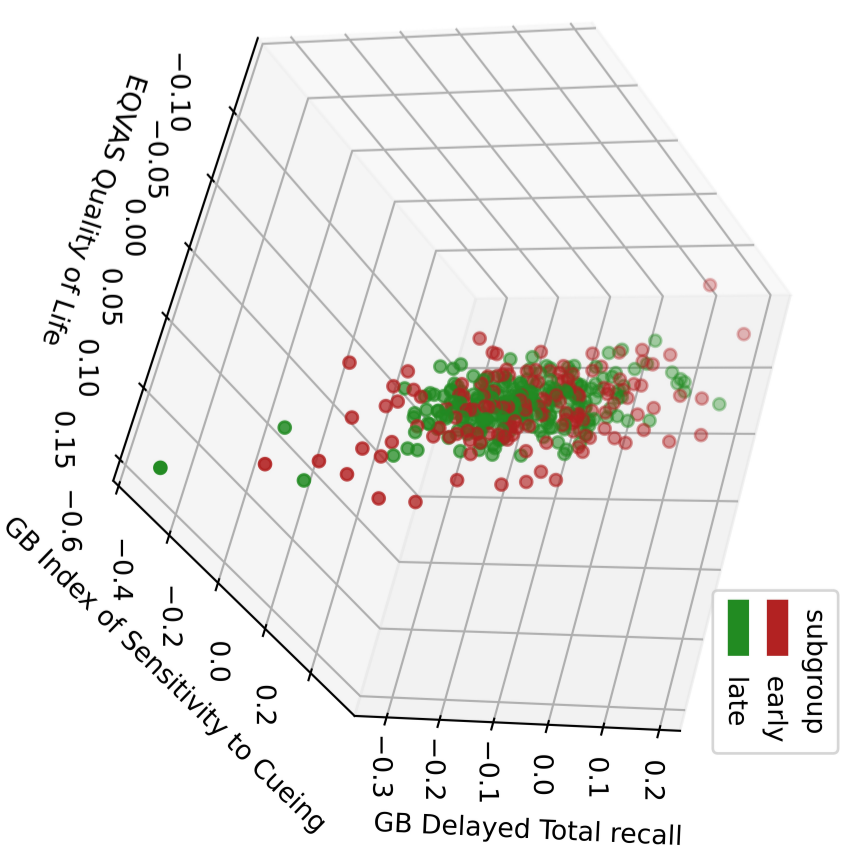
### A. Motor disability, Dependency & Focal Neurological Deficits



### B. Vadas-Cog subscores



### C. Life quality & Memory



## SUPPLEMENTARY MATERIAL

**Supplementary Table 1:** Percentages of visits with missing scores along a total of 2007 visits.

Score	Percentage of missing visits
Modified Rankin Scale	3%
Barthel Index	9%
NIHSS Index	10%
MDRS initiation	28%
EQVAS Quality of life	29%
TMT A Time	30%
GB Index of Sensitivity to Cueing	31%
GB Total Free Recall	31%
TMT B Time	35%
TMT B Errors	35%
VADAS – Cog Symbol Digit Test	42%
VADAS – Cog Backward Digit Span	42%
VADAS – Cog Digit Cancellation Task	42%
GB Delayed Total Recall	43%

**Supplementary Table 2:** Median (Q1, Q3) values of the different time shift ( $\tau$ ), progression rate ( $\xi$ ) and ordering of clinical score changes ( $\omega$ ) according to various baseline covariates (p values obtained with Mann Whitney U Tests)

Parameter	Gender		Education level		Smoking		Hypertension		Mutation location	
	male	female	low	high	yes	no	yes	No	EGFR 1-6	EGFR 7-34
$\tau$ (time shift)	<b>62.71</b> (54.56, 69.98)	<b>68.43</b> (62.22, 74.64)	<b>64.88</b> (57.45, 71.01)	<b>69.97</b> (62.99, 74.09)	<b>61.71</b> (52.07, 69.06)	<b>67.23</b> (59.97, 73.40)	<b>64.80</b> (59.96, 69.82)	<b>66.45</b> (59.59, 73.42)	<b>65.07</b> (57.70, 72.22)	<b>68.38</b> (61.69, 73.90)
p-value	<0.001		<0.001		<0.001		0.045		0.015	
$\xi$ (progression rate)	<b>0.10</b> (-0.33, 0.91)	<b>-0.07</b> (-0.40, 0.37)	0.02 (-0.35, 0.74)	0.03 (-0.39, 0.40)	-0.14 (-0.49, 0.42)	0.05 (-0.35, 0.61)	0.15 (-0.32, 0.84)	-0.02 (-0.39, 0.50)	0.05 (-0.36, 0.73)	0.00 (-0.36, 0.43)
p-value	0.013		0.401		0.118		0.107		0.429	
$\omega$ Barthel	0.001 (-0.03, 0.03)	0.009 (-0.02, 0.03)	<b>-0.001</b> (-0.03, 0.02)	<b>0.022</b> (0, 0.05)	0.015 (-0.02, 0.03)	0.006 (-0.03, 0.03)	-0.001 (-0.03, 0.02)	0.009 (-0.02, 0.03)	0.007 (-0.02, 0.03)	0.006 (-0.02, 0.03)
p-value	0.107		<0.001		0.308		0.053		0.798	
$\omega$ Rankin	0.004 (-0.06, 0.06)	0.019 (-0.03, 0.06)	<b>0.003</b> (-0.06, 0.05)	<b>0.042</b> (-0.01, 0.08)	0.024 (-0.04, 0.07)	0.009 (-0.04, 0.06)	<b>-0.003</b> (-0.06, 0.04)	<b>0.020</b> (-0.04, 0.06)	0.016 (-0.04, 0.06)	0.007 (-0.04, 0.05)
p-value	0.166		<0.001		0.329		0.026		0.564	
$\omega$ NIHSS	<b>-0.002</b> (-0.02, 0.01)	<b>0.003</b> (-0.01, 0.01)	<b>-0.002</b> (-0.01, 0.01)	<b>0.010</b> (0, 0.02)	0.001 (-0.01, 0.02)	0.001 (-0.01, 0.01)	<b>-0.004</b> (-0.01, 0.01)	<b>0.003</b> (-0.01, 0.02)	0.002 (-0.01, 0.01)	0 (-0.01, 0.01)
p-value	0.015		<0.001		0.961		0.021		0.610	
$\omega$ MDRS initiation	-0.004 (-0.08, 0.06)	-0.013 (-0.07, 0.05)	<b>0.001</b> (-0.06, 0.06)	<b>-0.031</b> (-0.11, 0.04)	-0.029 (-0.08, 0.05)	-0.004 (-0.07, 0.06)	-0.006 (-0.07, 0.05)	0.010 (-0.07, 0.06)	0 (-0.07, 0.06)	-0.014 (-0.07, 0.05)
p-value	0.488		<0.001		0.290		0.927		0.549	
$\omega$ TMT-A T	-0.007 (-0.09, 0.09)	-0.036 (-0.10, 0.06)	<b>-0.003</b> (-0.08, 0.09)	<b>-0.06</b> (-0.12, 0.02)	-0.054 (-0.11, 0.07)	-0.015 (-0.09, 0.08)	0.006 (-0.08, 0.10)	-0.029 (-0.10, 0.07)	-0.023 (-0.09, 0.08)	-0.012 (-0.09, 0.08)
p-value	0.282		<0.001		0.291		0.059		0.659	

<i>p-value</i>										
<b>ω</b> TMT-B T	-0.002 (- 0.16, 0.16)	-0.044 (- 0.15, 0.10)	<b>-0.002</b> (-0.14, 0.16)	<b>-0.103</b> (-0.21, 0.04)	-0.073 (-0.18, 0.10)	-0.021 (-0.15, 0.13)	0.011 (-0.14, 0.15)	-0.040 (-0.16, 0.12)	-0.027 (-0.15, 0.13)	-0.030 (-0.16, 0.13)
<i>p-value</i>	<i>0.260</i>		<i>&lt;0.001</i>		<i>0.231</i>		<i>0.113</i>		<i>0.935</i>	
<b>ω</b> TMT-B Errors	-0.001 (- 0.03, 0.03)	-0.005 (- 0.03, 0.02)	<b>0.002</b> (0.02, 0.03)	<b>-0.016</b> (-0.04, 0.01)	-0.009 (-0.03, 0.02)	-0.002 (-0.03, 0.02)	0.002 (-0.03, 0.02)	0.004 (-0.03, 0.02)	-0.003 (-0.03, 0.02)	-0.005 (-0.03, 0.02)
<i>p-value</i>	<i>0.327</i>		<i>&lt;0.001</i>		<i>0.186</i>		<i>0.498</i>		<i>0.715</i>	
<b>ω</b> EQVAS	<b>-0.001</b> (-0.02, 0.02)	<b>0.004</b> (0.01, 0.02)	0.001 (-0.02, 0.02)	0.004 (-0.01, 0.02)	0 (-0.02, 0.02)	0.002 (-0.02, 0.02)	0.001 (-0.02, 0.02)	0.002 (-0.02, 0.02)	0.002 (-0.02, 0.02)	0.002 (-0.02, 0.02)
<i>p-value</i>	<i>0.048</i>		<i>0.070</i>		<i>0.735</i>		<i>0.974</i>		<i>0.593</i>	
<b>ω</b> GBfree	<b>0.015</b> (0.09, 0.1)	<b>-0.016</b> (-0.09, 0.06)	<b>0.011</b> (0.07, 0.08)	<b>-0.059</b> (-0.12, 0.03)	-0.012 (-0.10, 0.08)	-0.006 (-0.09, 0.08)	<b>0.029</b> (-0.06, 0.08)	<b>-0.014</b> (-0.10, 0.07)	-0.010 (-0.09, 0.07)	0.004 (-0.10, 0.08)
<i>p-value</i>	<i>0.016</i>		<i>&lt;0.001</i>		<i>0.922</i>		<i>0.034</i>		<i>0.666</i>	
<b>ω</b> GBcueing	<b>0.017</b> (0.07, 0.10)	<b>-0.023</b> (-0.09, 0.05)	<b>-0.004</b> (-0.07, 0.08)	<b>-0.029</b> (-0.09, 0.05)	0.015 (-0.07, 0.09)	-0.016 (-0.08, 0.07)	0 (-0.06, 0.10)	-0.016 (-0.08, 0.06)	-0.007 (-0.08, 0.07)	-0.020 (-0.07, 0.07)
<i>p-value</i>	<i>0.003</i>		<i>0.041</i>		<i>0.315</i>		<i>0.113</i>		<i>0.989</i>	
<b>ω</b> GBdelayed	<b>0.011</b> (0.04, 0.06)	<b>-0.015</b> (-0.05, 0.03)	<b>0.001</b> (-0.04, 0.05)	<b>-0.020</b> (-0.05, 0.03)	0.005 (-0.04, 0.05)	-0.008 (-0.05, 0.04)	0.001 (-0.04, 0.06)	-0.009 (-0.05, 0.03)	-0.003 (-0.05, 0.04)	-0.011 (-0.04, 0.04)
<i>p-value</i>	<i>0.004</i>		<i>0.019</i>		<i>0.385</i>		<i>0.090</i>		<i>0.949</i>	
<b>ω</b> Backward-Digit	-0.002 (- 0.11, 0.12)	-0.021 (-0.11, 0.07)	<b>0.001</b> (0.09, 0.12)	<b>-0.050</b> (-0.17, 0.02)	-0.026 (-0.12, 0.07)	-0.006 (-0.11, 0.10)	0.008 (-0.10, 0.10)	-0.017 (-0.11, 0.10)	-0.015 (-0.11, 0.10)	-0.013 (-0.12, 0.10)
<i>p-value</i>	<i>0.334</i>		<i>&lt;0.001</i>		<i>0.175</i>		<i>0.284</i>		<i>0.873</i>	
<b>ω</b> Digit-Cancel	-0.017 (- 0.21, 0.16)	-0.061 (- 0.22, 0.13)	<b>-0.008</b> (-0.18, 0.17)	<b>-0.112</b> (-0.27, 0.07)	-0.072 (-0.26, 0.16)	-0.038 (-0.21, 0.14)	<b>0.015</b> (-0.15, 0.21)	<b>-0.064</b> (-0.23, 0.13)	-0.053 (-0.22, 0.13)	-0.013 (-0.21, 0.17)
<i>p-value</i>	<i>0.283</i>		<i>0.001</i>		<i>0.431</i>		<i>0.029</i>		<i>0.442</i>	
<b>ω</b> Symbol-Digit	-0.009 (- 0.16, 0.15)	-0.059 (- 0.16, 0.10)	<b>-0.006</b> (-0.13, 0.16)	<b>-0.101</b> (-0.21, 0.04)	-0.088 (-0.16, 0.10)	-0.027 (-0.16, 0.13)	0.011 (-0.13, 0.19)	-0.053 (-0.16, 0.11)	-0.042 (-0.16, 0.13)	-0.032 (-0.15, 0.13)
<i>p-value</i>	<i>0.284</i>		<i>&lt;0.001</i>		<i>0.272</i>		<i>0.057</i>		<i>0.692</i>	

**Legend:** The median values are presented with the first and third quartiles in brackets for each individual parameter according to the different baseline covariates. The p-values of Mann-Whitney U tests performed for each combination are given in italics. The distributions with a statistical difference are highlighted, and their median values are in bold.

**Supplementary Table 3:** Distribution of time shift ( $\tau$ ), progression rate ( $\xi$ ) and ordering of clinical score changes ( $\omega$ ) values in patients with early or late disease progression.

Parameter	Median (Q1,Q3)		p-value
	early	late	
Time shift ( $\tau$ )	59.02 (48.89 , 66.32)	69.24 (63.36 , 75.11)	4.91e-20
Progression rate ( $\xi$ )	0.84 (0.07 , 1.31)	-0.18 (-0.48 , 0.14)	6.50e-24
$\omega$ Barthel	-0.03 (-0.06 , 0.01)	0.02 (0 , 0.04)	3.15e-32
$\omega$ Rankin	-0.06 (-0.12 , -0.01)	0.04 (0 , 0.07)	1.26e-33
$\omega$ NIHSS	-0.01 (-0.02 , 0)	0.01 (0 , 0.02)	1.36e-13
$\omega$ MDRS-initiation	0.05 (-0.01 , 0.12)	-0.04 (-0.09 , 0.03)	6.13e-17
$\omega$ TMT-A T	0.1 (0.02 , 0.18)	-0.07 (-0.12 , -0.01)	3.26e-36
$\omega$ TMT-B T	0.18 (0.05 , 0.29)	-0.12 (-0.2 , -0.01)	2.62e-37
$\omega$ TMT-B E	0.03 (0 , 0.05)	-0.02 (-0.04 , 0)	3.50e-29
$\omega$ EQVAS	0 (-0.02 , 0.03)	0 (-0.01 , 0.02)	0.78
$\omega$ GB-free	0.06 (-0.03 , 0.14)	-0.04 (-0.11 , 0.03)	2.65e-13
$\omega$ GB-cueing	0 (-0.08 , 0.09)	-0.02 (-0.07 , 0.06)	0.31
$\omega$ GB-delayed	0 (-0.04 , 0.05)	-0.01 (-0.04 , 0.03)	0.07
$\omega$ Backward-Digit	0.12 (0.02 , 0.20)	-0.07 (-0.15 , 0)	8.56e-36
$\omega$ Digit-Cancel	0.18 (0.01 , 0.45)	-0.14 (-0.29 , 0)	1.20e-29
$\omega$ Symbol-Digit	0.17 (0.03 , 0.31)	-0.12 (-0.20 , -0.01)	2.13e-36

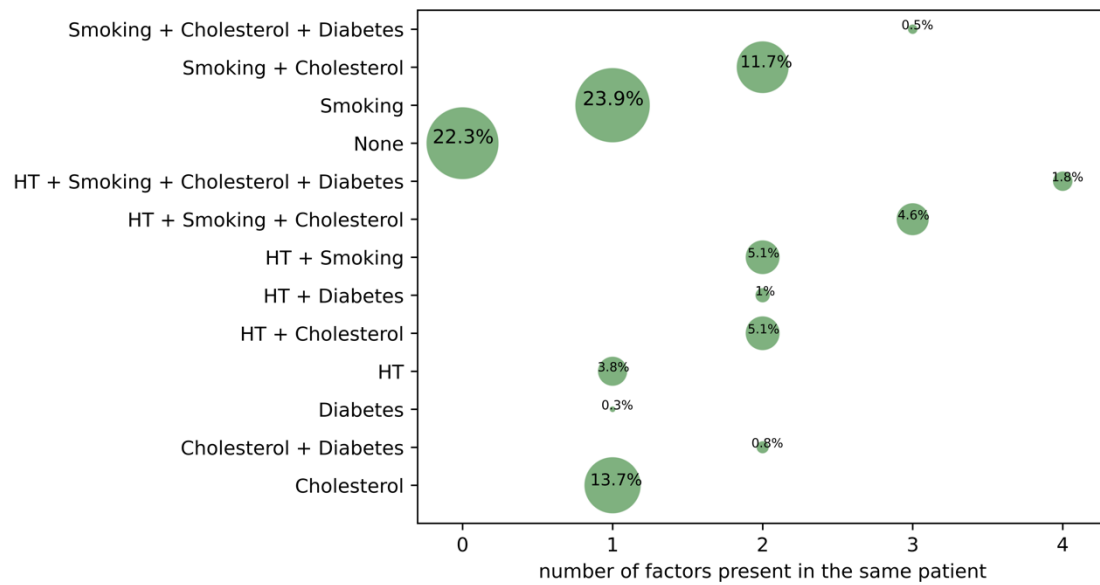
**Legend:** Median and the first and third quartile values of the different progression parameters are presented in the table. The p-values of the Mann-Whitney U test are given in the last column. The subgroup with late progression also presented with a slower progression rate. In this group, the Barthel, Rankin and NIHSS indices had a greater  $\omega$  values than in the early progression group indicating that motor disability, focal deficits and dependency develop later in these patients. In contrast, most  $\omega$  values for cognitive scores were smaller for the late subgroup indicating that the respective cognitive scores decreased earlier in this group of patients. The distribution of  $\omega$  values of the EQVAS score, and GB index of Sensitivity to Cueing and Delayed Total Recall did not differ between the groups with early or late disease progression.

**Supplementary Table 4: Associations between the different baseline covariates**

Covariate value (%)	Low education	Smoking	Hypertension	Hypercholesterolemia	Diabetes	Mutation location EGFR 1-6
Men (45%) <i>p-value</i>	45% <i>1.00</i>	52% <i>0.24</i>	49% <i>0.51</i>	49% <i>0.26</i>	55% <i>0.50</i>	49% <i>0.06</i>
Low education (72%) <i>p-value</i>		73% <i>0.90</i>	79% <i>0.09</i>	<b>80%</b> <b>0.006</b>	83% <i>0.40</i>	<b>68%</b> <b>0.04</b>
Smoking (18%) <i>p-value</i>			12% <i>0.14</i>	21% <i>0.37</i>	11% <i>0.64</i>	<b>22%</b> <b>0.003</b>
Hypertension (21%) <i>p-value</i>				<b>30%</b> <b>0.002</b>	<b>61%</b> <b>&lt;0.0001</b>	<b>16%</b> <b>0.0004</b>
Cholesterol (38%) <i>p-value</i>					<b>72%</b> <b>0.005</b>	37% <i>0.58</i>
Diabetes (5%) <i>p-value</i>						4% <i>0.88</i>

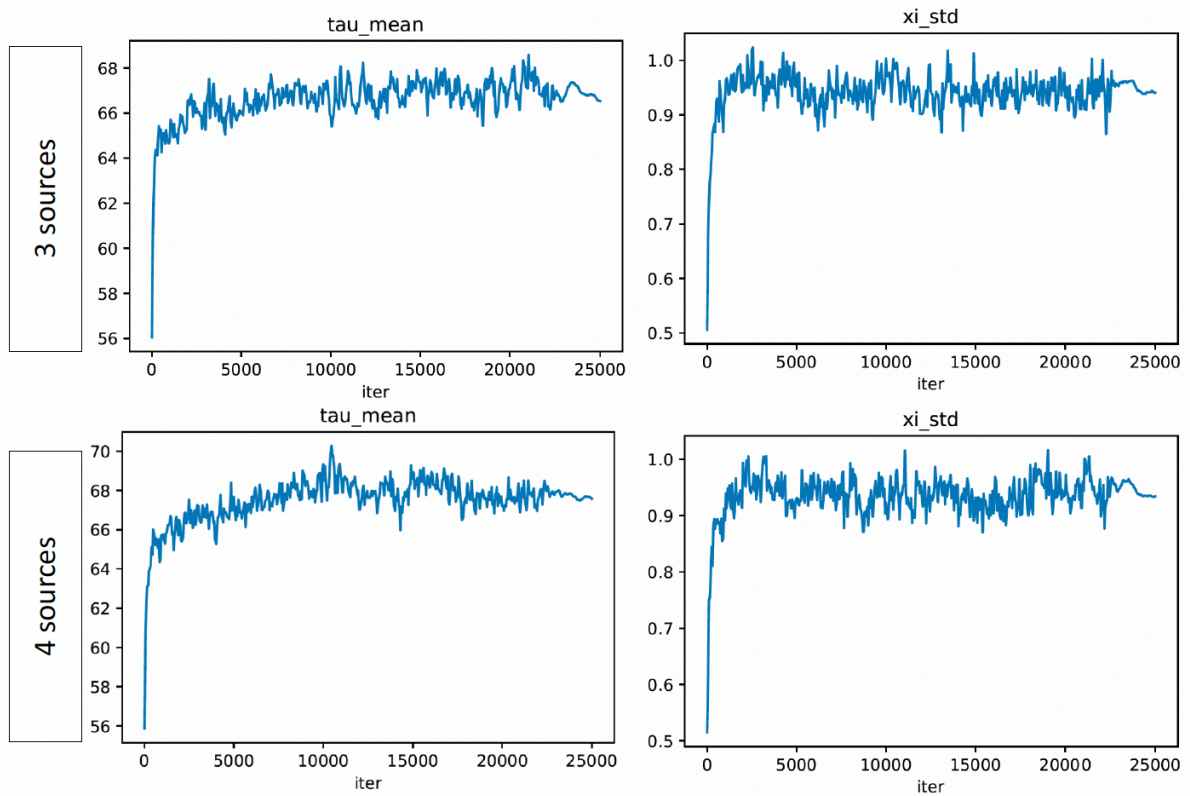
**Legend:** The associations between the baseline covariates are assessed with chi-2 tests. The observed relative frequencies of each covariate are given for the whole population in the first column and then for each combination of covariates in the following columns. The p-values are given for each pair of covariates in italics. The statistically significant associations are in bold. A low education level is defined by less than 13 years of education.

**Supplementary Figure 1: Frequency of cardiovascular risk factors in the CADASIL population**



**Legend:** The relative frequency of each cardiovascular risk factor and of their combinations in the study sample are presented on this figure. The y axis corresponds to the cardiovascular risk factors and the x axis to the number of factors coexisting in the same patient. There were 5.6% missing data (HT = Hypertension)

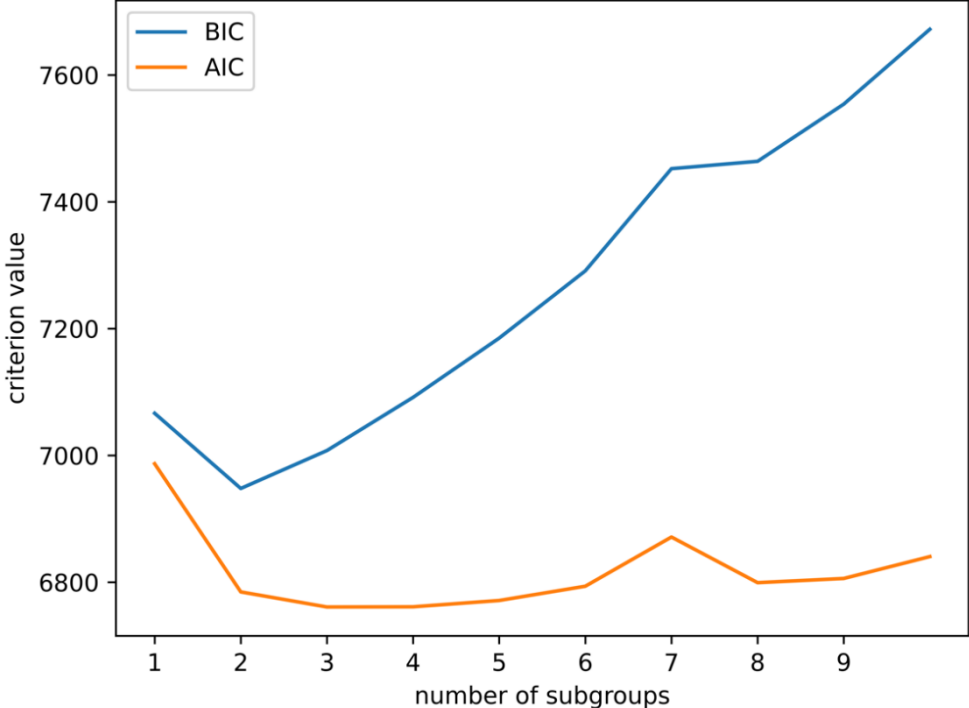
**Supplementary Figure 2: Convergence of the population parameters**



**Legend:** Convergence of the population parameters over 2500 iterations of the algorithm. Two Leaspy models were tested using three (first line) and 4 sources (second line). The mean of the time shift ( $\tau$ ) and the standard deviation of the acceleration rate ( $\xi$ ) are presented in the left and right column respectively.

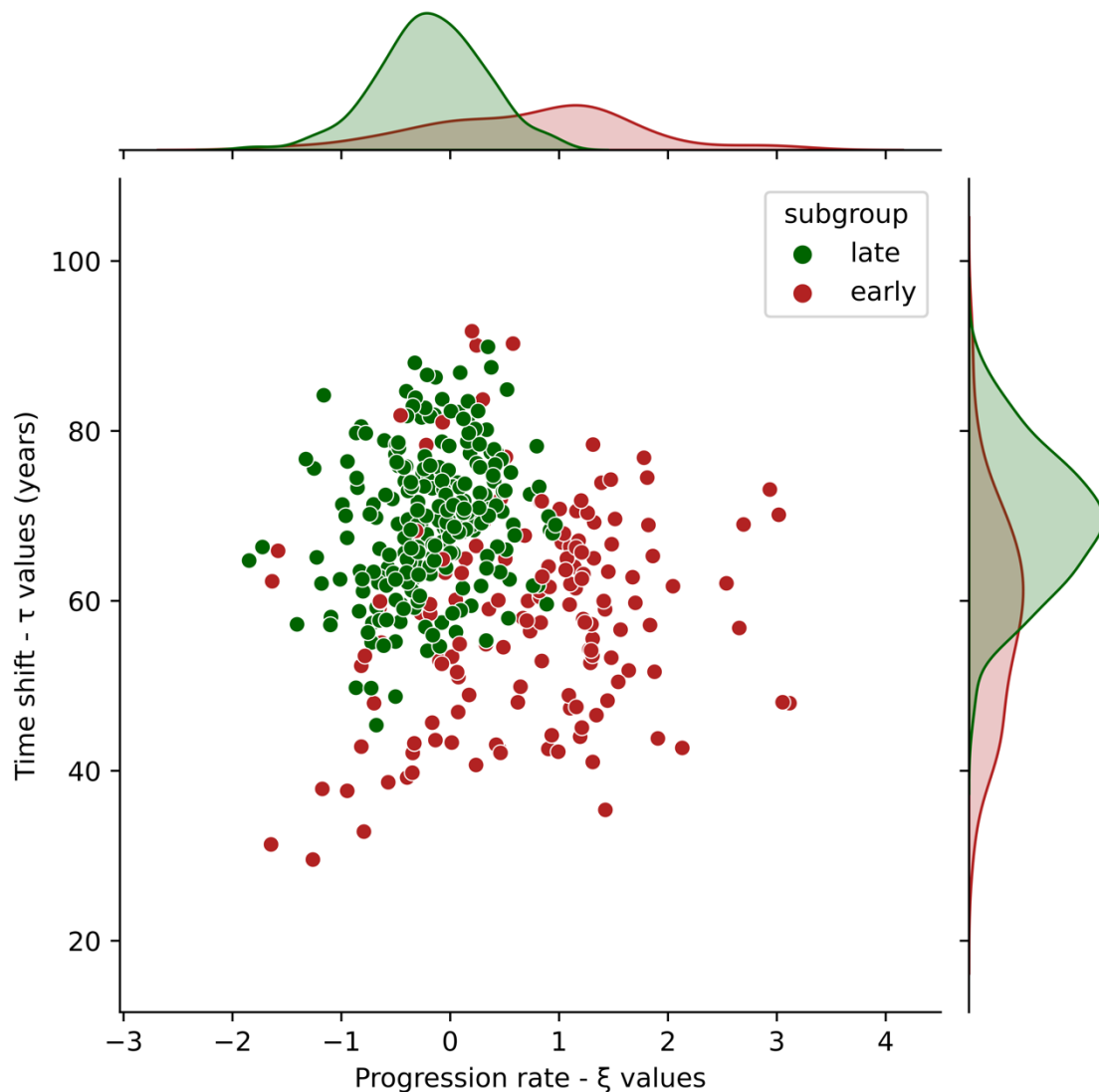


**Supplementary Figure 3:** Optimization criteria for the GMM clustering algorithm



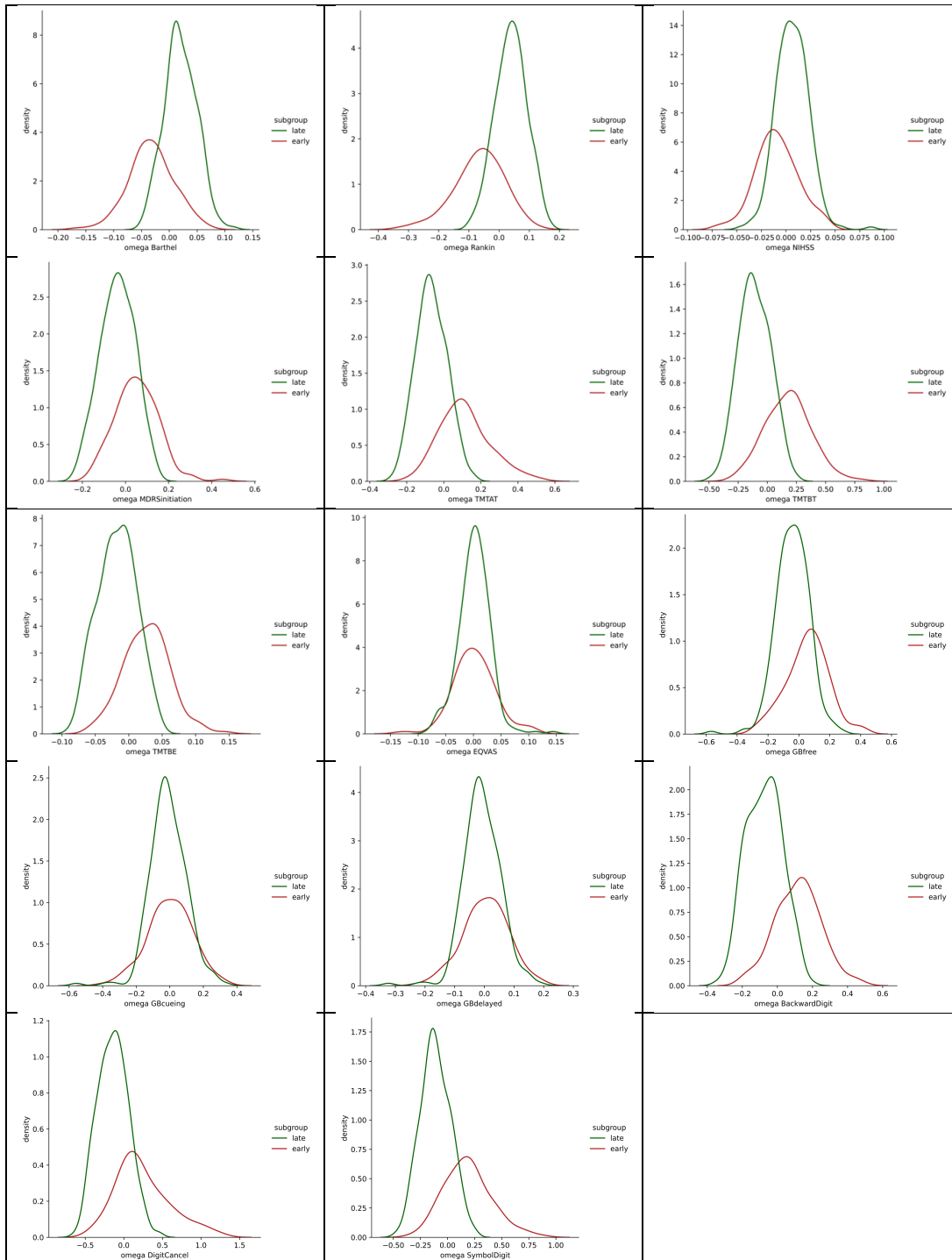
**Legend:** AIC and BIC values obtained by the GMM clustering algorithm according to the possible number of subgroups. A smaller value for these criteria indicates a better performing model.

**Supplementary Figure 4:** Distribution of time shift ( $\tau$  values) and progression rate ( $\xi$  values) at individual level revealing two groups of patients with distinct clinical progression types



**Legend:** Each point corresponds to one patient. The y-axis corresponds to the time-shift  $\tau$  in years. The group with small  $\tau$  values corresponds to patients with an earlier clinical progression (and presumably with an earlier age of onset) compared to patients with high  $\tau$  values. The x-axis represents the progression rate ( $\xi$  values).  $\xi$  values  $> 0$  indicate a faster disease progression than the average observed in the whole sample. The green points correspond to patients with a relatively late and slow clinical progression. The red points correspond to patients with a relatively early and fast clinical progression. On the upper and the right axis, the distributions of  $\tau$  and  $\xi$  values also shown.

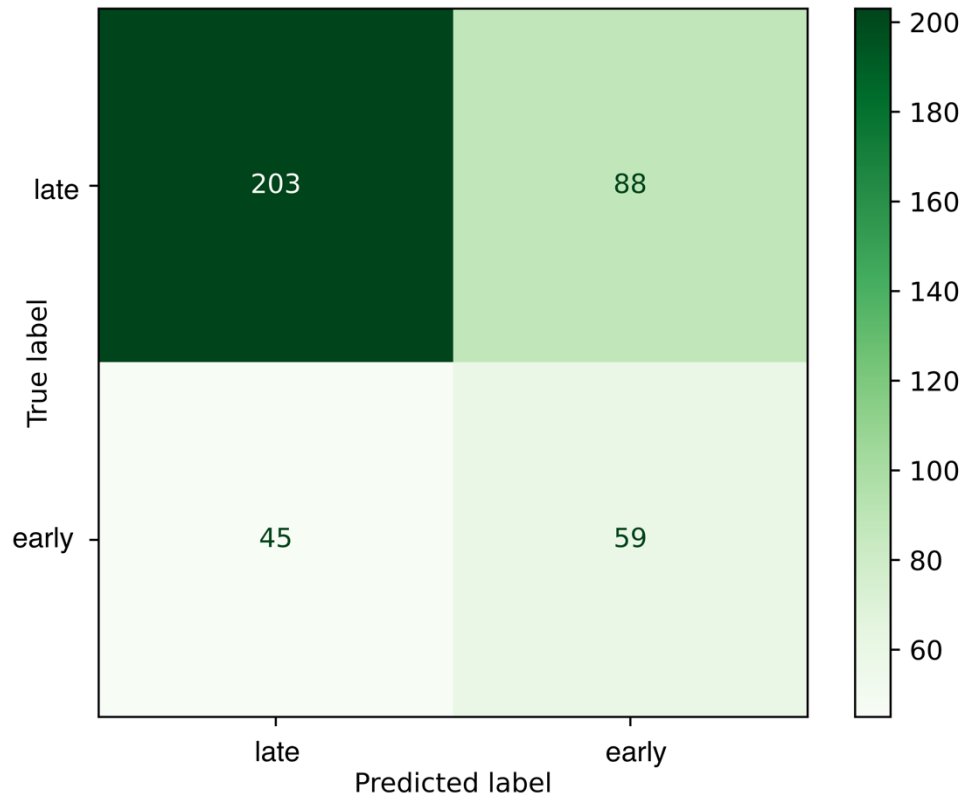
**Supplementary Figure 5: Subgroup-specific distributions of the intermarker spacing parameters ( $\omega$ )**



**Legend:** The distribution for the intermarker spacing parameters (omegas) are shown for each subgroup. A greater intermarker spacing for one subgroup suggests that the respective score starts its deterioration later than for the other subgroup. The green lines correspond to the subgroup that shows a late age of onset, slow progression rate and shows earlier symptoms of cognitive decline. The red lines correspond to the subgroup that shows an early age of onset or a fast rate of progression and shows earlier symptoms of motor disability and neurological deficits.

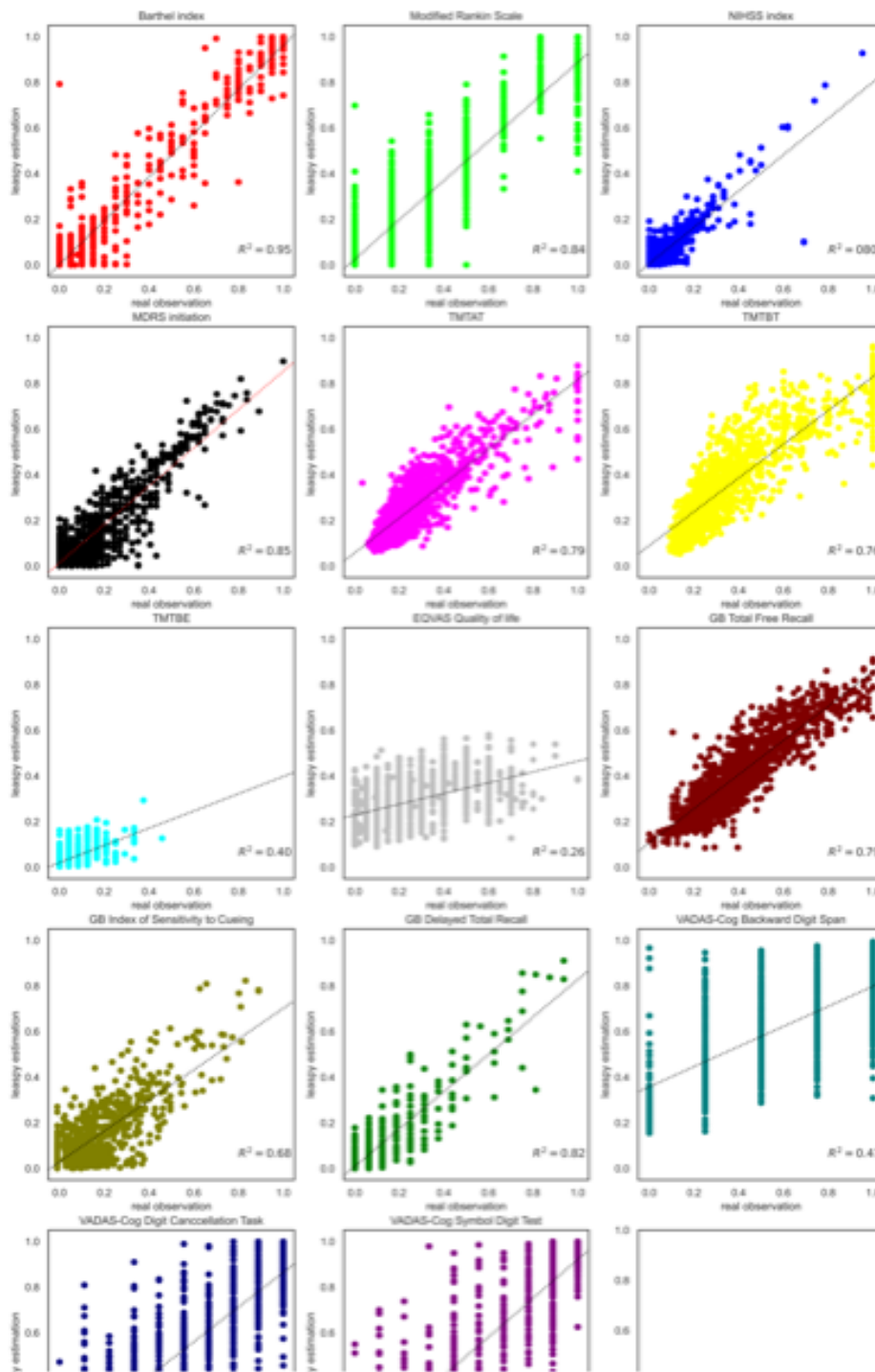
**Supplementary Figure 6 :** Confusion matrix for the logistic regression predicting the subgroup label.

Confusion Matrix for the Multivariable Logistic Regression Model



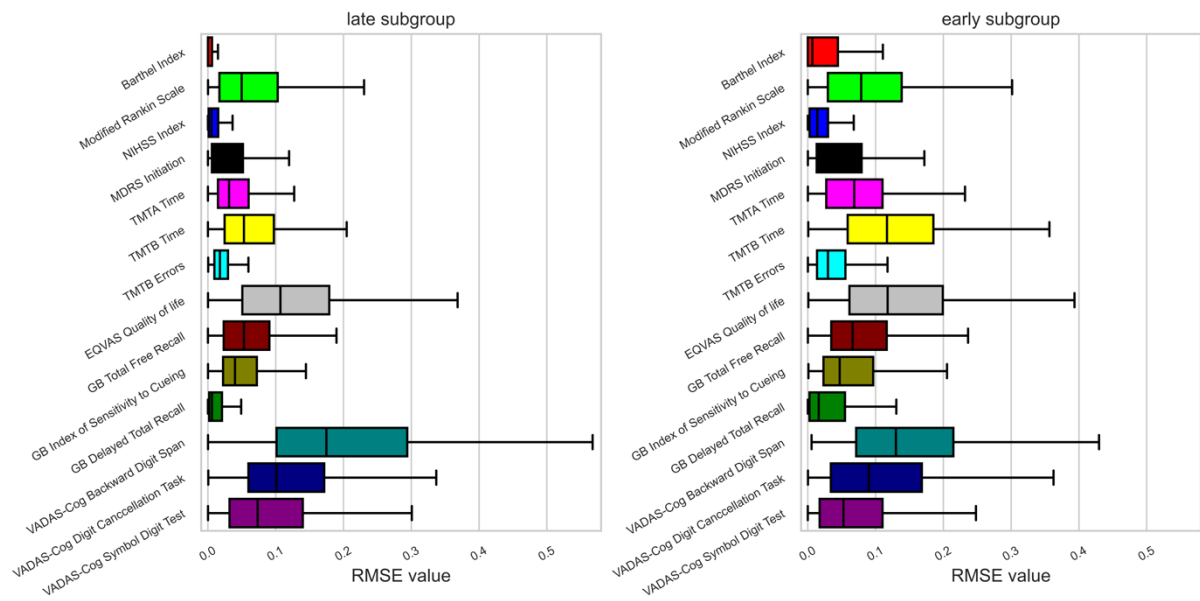
**Legend:** Confusion matrix showing the amount of correct predictions the logistic regression is able to produce, using as a dependent variable the subgroup label, and as independent variables the gender, the education level, the hypertension and the location of the pathogenic variant.

**Supplementary Figure 7 : Reconstruction error.**



**Legend:** Scatter plots comparing the estimated vs the real values for each score. The dotted lines correspond to the regression lines. The  $R^2$  coefficients are stated in the bottom right corner of each plot.

### Supplementary Figure 8 : Root Mean Square Error par group.



**Legend:** Boxplots comparing the distributions of Root Mean Square Error values for the 14 scores between the two groups.