eMethods

Steps of the model building process

Latent process mixed models were used to model the trajectory of the 16 scores selected for analysis[1,2]. These models can be considered extensions of standard linear mixed models as they involve a latent process corresponding to the actual quantity of interest for which test scores represent imprecise quantitative measures. These models are divided into two parts: (1) a linear mixed model, which describes the latent process according to time and potential covariates, and (2) a link function, which relates the observations with the latent process.

The first step was to select the best parameterized function for linking each score to its underlying latent process. For this purpose, we estimated latent process mixed models assuming a similar trajectory of the underlying latent process with a different link function for each model. The change over time of the latent process underlying each score was first described using a cubic function of time with correlated random effects on the intercept, slope, quadratic slope, and cubic slope. We compared eight link functions according to Akaike information criteria (AIC): linear transformation, beta cumulative distribution function (CDF), and quadratic I-splines with three, five, or seven nodes located at the quantiles of the test distribution or equidistant. The models were not adjusted for any covariate since we aimed to analyse the relationship between each score and its underlying latent process.

In a second step, we ascertained the shape of the trajectory of each score using the previously selected transformation. To test whether the addition of a random effect improved the model fit, we compared nested models using an approximation of the distribution of the likelihood ratio statistic as proposed by Stram and Lee.[3] At minimum, a random effect on the linear slope was kept in the model. We used AIC to select the best variance-covariance structure of the random effects between unstructured and diagonal and to ascertain the need for an autocorrelated Gaussian process. To test whether an additional fixed effect was different from 0, we compared nested models using the likelihood ratio test (LRT).

Finally, assuming that the population was heterogeneous and composed of *G* latent classes of subjects characterized by *G* mean profiles of trajectories, we used latent class mixed models, which are divided into two parts [4-6]: (1) the probability of belonging to a latent class modelled using a multinomial logistic regression according to all information collected, and (2) a class-specific trajectory, which is described by a latent process mixed model. The different models obtained with one to three latent classes were compared according to the Bayesian Information Criterion (BIC) and the Integrated Classification Likelihood Criterion with BIC approximation (ICL-BIC).[7]

References

[1] Brice S, Jabouley A, Reyes S, Machado C, Rogan C, Dias-Gastellier N, et al. Modeling the Cognitive Trajectory in CADASIL. Journal of Alzheimer's disease : JAD. 2020;77:291-300.

[2] Proust C, Jacqmin-Gadda H, Taylor JM, Ganiayre J, Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. Biometrics. 2006;62:1014-24.

[3] Stram D, Lee J. Variance Components Testing in the Longitudinal Mixed Effects Model. Biometrics. 1994;50:1171.

[4] Muthen B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics. 1999;55:463-9.

[5] Lin H, Turnbull BW, Mc Culloch CE, EH S. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. Journal of the American Statistical Association 2002;97:53–65.

[6] Lin H, McCulloch CE, Turnbull BW, Slate EH, Clark LC. A latent class mixed model for analysing biomarker trajectories with irregularly scheduled observations. Stat Med. 2000;19:1303-18.

[7] Han J, Slate EH, Pena EA. Parametric latent class joint model for a longitudinal biomarker and recurrent events. Stat Med. 2007;26:5285-302.

eTable 1. Numbers of observations and patients per 5-year age range

Age range	No. of observations	No. of patients
[25.3-30.3[	23	11
[30.3-35.3[	65	27
[35.3-40.3[	80	40
[40.3-45.3[	133	56
[45.3-50.3[	176	76
[50.3-55.3[	203	85
[55.3-60.3[	200	87
[60.3-65.3[	170	71
[65.3-70.3[	138	58
[70.3-75.3[	55	28

eTable 2. Link functions of the best selected latent process models

Variable	Link function
Total Free Recall	Quadratic I-splines with 3 equidistant nodes*
Index of Sensitivity to Cueing	Quadratic I-splines with 3 nodes at quantiles†
Delayed Total Recall	Quadratic I-splines with 7 equidistant nodes‡
Digit Cancellation Test	Quadratic I-splines with 5 equidistant nodes§
Symbol Digit Test	Quadratic I-splines with 7 equidistant nodes‡
Backward Digit Span	Quadratic I-splines with 5 equidistant nodes§
Initiation/Perseveration	Quadratic I-splines with 7 equidistant nodes‡
TMT A Time	Quadratic I-splines with 7 nodes at quantiles‖
TMT B Errors	Beta CDF #
TMT B Time	Quadratic I-splines with 7 nodes at quantiles‖
MDRS Total score	Quadratic I-splines with 5 nodes at quantiles**
VADAS-Cog Total score	Beta CDF
NIHSS	Beta CDF
Barthel index	Quadratic I-splines with 7 equidistant nodes‡
Modified Rankin scale	Quadratic I-splines with 7 equidistant nodes‡
EQ VAS	Beta CDF

*Three equidistant nodes corresponding to Total Free Recall scores 1.00, 24.50 and 48.00.

†Three nodes located at 0th, 50th, and 100th quantiles corresponding to Index of Sensitivity to Cueing scores 22.50, 95.83, and 100.00.

‡Seven equidistant nodes corresponding to (1) Delayed Total Recall scores 3.00, 5.17, 7.33, 9.50, 11.67, 13.83, and 16.00; (2) Symbol Digit Correct Numbers scores 1.00, 2.50, 4.00, 5.50, 7.00, 8.50, and 10.00; (3) Initiation/Perseveration subscores 4.00, 9.50, 15.00, 20.50, 26.00, 31.50, and 37.00; (4) Barthel index scores 0.00, 16.67, 33.33, 50.00, 66.67, 83.33, and 100.00; and (5) Modified Rankin scale scores 0.00, 0.83, 1.67, 2.50, 3.33, 4.17, and 5.00.

§Five equidistant nodes corresponding to (1) Digit Cancellation Task (correct numbers crossed off) scores 1.00, 3.25, 5.50, 7.75, and 10.00; and (2) Backward Digit Span scores 1.00, 2.00, 3.00, 4.00, and 5.00.

‖Seven nodes located at 0th, 14.29th, 28.57th, 42.86th, 57.14th, 71.43th, 85.71th, and 100th quantiles corresponding to (1) TMT A Time scores 6.00, 23.00, 30.00, 37.00, 46.00, 60.00, and 308.00; and (2) TMT B Time scores 27.00, 53.00, 69.00, 87.50, 115.00, 180.00, and 618.00.

#CDF: cumulative distribution function

**Five nodes located at 0th, 25th, 50th, 75th, and 100th quantiles corresponding to MDRS Total scores 66.00, 136.00, 141.00, 143.00, and 144.00.

eTable 3. Selection of the best univariate models in the latent process scale.

Variable	Fixed effect*	Random effect†	Matrix of variance-covariance	Autocorrelated process	Latent classes
Total Free Recall‡	Quadratic	Quadratic	Unstructured	Brownian motion	1
Index of Sensitivity to Cueing‡	Quadratic	Quadratic	Unstructured	None	2
Delayed Total Recall‡	Quadratic	Linear	Diagonal	None	2
Digit cancellation test‡	Cubic	Quadratic	Unstructured	Brownian motion	1
Symbol digit test‡	Linear	Linear	Diagonal	Brownian motion	1
Backward digit span‡	Linear	Linear	Unstructured	None	1
Initiation/Perseveration‡	Quadratic	Quadratic	Unstructured	None	2
TMT A Time‡	Quadratic	Quadratic	Unstructured	None	1
TMT B Errors‡	Linear	Linear	Diagonal	None	1
TMT B Time§	Quadratic	Linear	Diagonal	Autoregressive process	1
MDRS Total score‡	Quadratic	Quadratic	Unstructured	Brownian motion	1
VADAS-Cog Total score‡	Quadratic	Quadratic	Unstructured	None	1
NIHSS	Linear	Linear	Diagonal	Autoregressive process	1
Barthel index	Quadratic	Quadratic	Unstructured	None	2
Modified Rankin scale	Quadratic	Quadratic	Unstructured	Brownian motion	1
EQ VAS	Linear	Linear	Diagonal	Autoregressive process	1

*The shape of the trajectory resulted in a Linear, Quadratic, or Cubic function of time.

†The model included random effects on the intercept and slope (Linear) or random effects on the intercept, slope, and quadratic slope (Quadratic).

‡The model was adjusted for education.

§The model was adjusted for education and includes the interaction between education and time, as the latter was significant (p = 0.002). The p-value is derived from the likelihood ratio test involving the model that includes education at baseline versus the model that includes education at baseline and the interaction between education and time.

To link each score with its underlying latent process, we used the link functions displayed in Supplementary Table 2.

eAppendix 1. Results of latent class mixed models

Index of Sensitivity to Cueing

(i) Selection of the optimal number of latent classes

G	npm	loglik	BIC	ICL-BIC	Entropy	Frequency of latent classes (%)		
						1	2	3
1	14	-2775.93	5629.98			100.0		
2	**19**	**-2756.50**	**5619.02**	**5722.57**	**0.72**	**82.6**	**17.4**	
3	24	-2747.00	5627.92	5894.83	0.54	16.2	64.9	18.9

(ii) Posterior classification

Final classification	Number of subjects (%)	Mean posterior probability to belong to class (%)	
		Early decline	Late decline
Group 1: Early decline	46 (17.4%)	88.8	11.2
Group 2: Late decline	219 (82.6%)	7.5	92.5

(iii) Posterior probabilities above a threshold (%)

Threshold	Early decline	Late decline
> 0.7	82.6	92.2
> 0.8	76.1	85.4
> 0.9	67.4	74.4

Delayed Total Recall

(i) Selection of the optimal number of latent classes

G	npm	loglik	BIC	ICL-BIC	Entropy	Frequency of latent classes (%)		
						1	2	3
1	14	-580.16	1238.43			100.0		
2	**20**	**-504.00**	**1119.59**	**1147.79**	**0.92**	**20.4**	**79.6**	
3	26	-477.98	1101.02	1199.93	0.83	14.3	70.6	15.1

(ii) Posterior classification

Final classification	No. of subjects (%)	Mean posterior probability to belong to class (%)	
		Decline	No decline
Group 1: Early decline	54 (20.4%)	94.8	5.2
Group 2: No decline	211 (79.6%)	1.4	98.6

(iii) Posterior probabilities above a threshold (%)

Threshold	Early decline	Late decline
> 0.7	94.4	98.1
> 0.8	90.7	96.2
> 0.9	83.3	95.7

Initiation/Perseveration subscore of MDRS

(i) Selection of the optimal number of latent classes

G	npm	loglik	BIC	ICL-BIC	Entropy	Frequency of latent classes (%)		
						1	2	3
1	18	-1843.21	3786.85			100.0		
2	**24**	**-1806.26**	**3746.44**	**3826.20**	**0.78**	**77.4**	**22.6**	
3	30	-1787.35	3742.10	3907.12	0.72	16.2	69.4	14.3

(ii) Posterior classification

Final classification	No. of subjects (%)	Mean posterior probability to belong to class (%)	
		Early decline	Late decline
Group 1: Early decline	60 (22.6%)	90.1	9.9
Group 2: Late decline	205 (77.4%)	4.5	95.5

(iii) Posterior probabilities above a threshold (%)

Threshold	Early decline	Late decline
> 0.7	83.3	96.6
> 0.8	78.3	95.6
> 0.9	73.3	82.9

Barthel index

(i) Selection of the optimal number of latent classes

G	npm	loglik	BIC	ICL-BIC	Entropy	Frequency of latent classes (%)		
						1	2	3
1	17	-2468.92	5032.70			100.0		
2	**21**	**-2400.91**	**4918.99**	**4944.66**	**0.93**	**8.3**	**91.7**	
3	25	-2366.56	4872.62	4953.67	0.86	9.1	86.4	4.5

(ii) Posterior classification

Final classification	No. of subjects (%)	Mean posterior probability to belong to class (%)	
		Intermediate decline	Late decline
Group 1: Intermediate decline	22 (8.3%)	93.4	6.6
Group 2: Late decline	243 (91.7%)	1.2	98.8

(iii) Posterior probabilities above a threshold (%)

Threshold	Intermediate decline	Late decline
> 0.7	86.4	99.6
> 0.8	86.4	99.2
> 0.9	86.4	97.9

For each latent class mixed model, the two latent classes provided very good discrimination, with entropy measures ranging from 0.72 to 0.93 and proportions of maximal posterior probabilities above 0.8 at least equal to 76.1% in classes 1 to 2. Mean maximal posterior probabilities of subjects classified in each latent class were remarkably close to 1 (i.e., all $\geq$ 88.8%).

eTable 4. Baseline characteristics predicting an earlier or larger decline for 4 scores (Index of Sensitivity to Cueing, Delayed Total Recall, Initiation/perseveration subscore, Barthel Index) in a subgroup of patients (univariate logistic regressions with multiple imputation of missing data)

		Index of Sensitivity to Cueing		Delayed Total Recall		Initiation/Perseveration subscore		Barthel index	
Baseline characteristic	% NA*	OR† 95% CI‡	*p*-value	OR 95% CI	*p*-value	OR 95% CI	*p*-value	OR 95% CI	*p*-value
Age (years)	0.0	0.99 [0.96; 1.02]	0.417	**1.05 [1.02-1.08]**	**0.001**	1.00 [0.98-1.03]	0.834	1.04 [1.00-1.08]	0.086
Sex: Male	**0.0**	**4.38 [2.15; 8.93]**	**< 0.001**	**3.74 [1.96-7.13]**	**< 0.001**	**2.59 [1.43-4.70]**	**0.002**	1.50 [0.62-3.60]	0.365
Education: > high school diploma	0.0	0.81 [0.42; 1.53]	0.507	**0.38 [0.20-0.72]**	**0.003**	**0.31 [0.16-0.58]**	**< 0.001**	**0.22 [0.07-0.66]**	**0.007**
Smoking									
Never		1.00 -		1.00 -		1.00 -		1.00 -	
Former	15.5	0.70 [0.30; 1.62]	0.404	0.73 [0.34; 1.56]	0.416	1.04 [0.52; 2.11]	0.903	0.59 [0.22; 1.6]	0.301
Current		1.73 [0.78; 3.85]	0.176	1.11 [0.5; 2.43]	0.803	1.35 [0.61; 2.99]	0.454	0.48 [0.14; 1.67]	0.248
Alcohol consumption									
Never		1.00 -		1.00 -		1.00 -		1.00 -	
< 2 glasses of wine per day (male)	16.6	1.61 [0.86; 1.88]	0.228	1.18 [0.58; 2.40]	0.649	1.29 [0.66; 2.54]	0.458	1.71 [0.53; 5.5]	0.366
> 2 glasses of wine per day (male)		1.32 [0.47; 3.65]	0.597	2.06 [0.77; 5.53]	0.152	1.43 [0.55; 3.74]	0.462	**3.13 [1.02; 9.63]**	0.047
Hypertension: Yes	15.5	1.94 [0.96; 3.94]	0.066	**2.17 [1.12; 4.23]**	**0.022**	1.84 [0.93; 3.64]	0.08	2.49 [0.99; 6.23]	0.051
Diabetes: Yes	15.5	1.39 [0.38; 5.09]	0.622	1.23 [0.35; 4.36]	0.751	0.91 [0.22; 3.83]	0.897	1.93 [0.41; 9.09]	0.404
Systolic blood pressure (mmHg)	1.9	1.00 [0.98; 1.02]	0.703	1.01 [0.99; 1.03]	0.207	1.01 [0.99; 1.03]	0.215	**1.03 [1; 1.05]**	**0.034**
Diastolic blood pressure (mmHg)	1.9	1.01 [0.98; 1.04]	0.609	1.01 [0.99; 1.04]	0.288	1.02 [1; 1.05]	0.095	1.04 [1; 1.08]	0.078
Homocysteine (µmol/L)	16.2	1.04 [0.98; 1.11]	0.178	1.09 [1.03; 1.16]	0.005	1.05 [0.99; 1.1]	0.102	1.01 [0.94; 1.1]	0.742
Previous stroke events: Yes	0.0	1.92 [1.00; 3.70]	0.051	1.61 [0.88-2.96]	0.122	1.86 [1.03-3.35]	0.038	**2.92 [1.11-7.71]**	**0.031**
Gait disturbances: Yes	0.0	1.32 [0.59; 2.99]	0.500	**4.81 [2.38-9.70]**	**< 0.001**	**3.49 [1.75-6.97]**	**< 0.001**	**28.38 [9.67-83.30]**	**< 0.001**
Balance problems: Yes	0.0	0.94 [0.42; 2.09]	0.880	1.70 [0.85-3.38]	0.133	1.60 [0.82-3.13]	0.171	**5.74 [2.33-14.16]**	**< 0.001**
Disability: Moderate or severe	0.4	2.31 [0.68; 7.87]	0.180	**6.95 [2.18; 22.22]**	**0.001**	**12.85 [3.41; 48.38]**	**< 0.001**	**66.37 [16.13-273.03]**	**< 0.001**
Dementia: Yes	0.0	2.44 [0.43; 13.75]	0.311	4.08 [0.80-20.80]	0.091	**18.54 [2.12-162.03]**	**0.008**	**63.6 [15.48; 261.32]**	**< 0.001**
MADRS	1.9	1.02 [0.94; 1.11]	0.609	11.02 [0.98; 1.06]	0.372	**1.05 [1.01; 1.09]**	**0.014**	1.03 [0.98; 1.09]	0.212

**NA: number of missing data that were imputed; †OR: odds ratio; ‡CI: confidence intervalThe modelled probability is that of belonging to the most severely affected group of patients (i.e., "Early decline" for Index of Sensitivity to Cueing and Initiation/Perseveration, "Decline" for Delayed Total Recall, and "Intermediate decline" for Barthel index).*