Prediction of 3-year clinical course in CADASIL

Eric Jouvent, Edouard Duchesnay, Foued Hadj-Selem, François de Guio,
Jean-François Mangin, Dominique Hervé, Marco Duering, Stefan Ropele,
Reinhold Schmidt, Martin Dichgans, et al.

▶ **To cite this version:**

Prediction of 3-year clinical course in CADASIL

Eric Jouvent, MD, PhD*
Edouard Duchesnay, PhD*
Foued Hadj-Selem, PhD
François De Guio, PhD
Jean-François Mangin, PhD
Dominique Hervé, MD
Marco Duering, MD
Stefan Ropele, PhD
Reinhold Schmidt, MD
Martin Dichgans, MD
Hugues Chabriat, MD, PhD

Correspondence to
Dr. Chabriat:
hugues.chabriat@aphp.fr

ABSTRACT

Objective: To obtain simple models predicting disease evolution at 3 years for a given patient with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL).

Methods: Based on data obtained in a prospective study of 236 patients, we built and validated models predicting, at the individual level, 3-year changes in Mini-Mental State Examination (MMSE), Mattis Dementia Rating Scale (MDRS), Trail Making Test version B (TMTB), and modified Rankin Scale (mRS). These models were based on different sets of predictors obtained at baseline, including either clinical data (epidemiologic data and cardiovascular risk factors) or clinical data and quantitative MRI markers (volume of lacunes [LL_V], volume of white matter hyperintensities, normalized brain volume [BPF], number of microbleeds). The Bayesian information criterion (BIC) and the coefficient of determination (R^2) were used to determine models with the highest predictive ability and the lowest numbers of predictors.

Results: We obtained validated models with a demonstrated ability to predict, for a given patient, 3-year changes in MMSE, MDRS, TMTB, and mRS (R^2 on independent samples: 0.22, 0.12, 0.09, and 0.17, respectively). In all cases, the best models according to R^2 and BIC values included only the baseline values of the outcome, of BPF, and of LL_V. Inclusion of other potential predictors always led to a loss of generalizability.

Conclusions: The prediction of 3-year changes in MMSE, MDRS, TMTB, and mRS for a given patient with CADASIL can be obtained using simple models relying only on the initial values of the considered score, BPF, and LL_V. *Neurology®* 2016;87:1787–1795

GLOSSARY

BIC = Bayesian Information Criterion; **BPF** = brain parenchymal fraction; **CADASIL** = cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; **CI** = confidence interval; **LLV** = volume of lacunes; **MBN** = number of microbleeds; **MDRS** = Mattis Dementia Rating Scale; **MMSE** = Mini-Mental State Examination; **mRS** = modified Rankin Scale; **MSPE** = mean squared prediction error; **SVD** = small vessel disease; **TMTB** = Trail Making Test part B; **WMHV** = white matter hyperintensities of presumed vascular origin.

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is the most frequent hereditary cerebral small vessel disease (SVD), caused by mutations of the *NOTCH3* gene.[1] In CADASIL, the clinical course appears highly variable. Some patients develop dementia and become heavily disabled around their 50s while others live independently during their 80s.[2] Age, male sex, active smoking, brain atrophy, and volume of lacunes are associated with clinical severity and with further clinical worsening,[3–5] while associations between other factors, such as blood pressure, number of microbleeds, or volume of white matter hyperintensities, and clinical severity have been inconsistently observed.[6,7] Among all these factors, those that actually predict disease evolution for a given patient remain undetermined. Identifying such factors would be useful for patient care and for selection of homogeneous samples in clinical trials.

We hypothesized that in CADASIL, predictive models may help determine whether a given patient is at high risk to deteriorate or, by contrast, is likely to remain stable. For this purpose, we aimed to build, validate, and compare different models based on the smallest number of clinical variables and quantitative MRI markers to predict how cognitive and disability scores of a given patient will change over the next 3 years.

METHODS **Participants.** A total of 290 patients older than 18 years were recruited in a cohort study of patients with CADASIL at Lariboisière (Paris) or Ludwig-Maximilians-Universität (Munich) hospitals, the 2 main referral centers for CADASIL in France and Germany, between October 2003 and September 2010. All participants harbored a typical mutation of the *NOTCH3* gene. Of these 290 patients, 236 patients had available data both at baseline and after 3 years of follow-up (14 died, 23 did not come back because of too severe disability or dementia, 9 refused to attend the 3-year evaluation, 3 were lost to follow-up, and 5 more were not evaluated for other reasons). As previously reported, clinical, biological, and MRI data were obtained every 18 months during follow-up.[4,8] Global cognitive performances were assessed by the Mini-Mental State Examination (MMSE) and the Mattis Dementia Rating Scale (MDRS), which includes a brief evaluation of executive functions. Processing speed was assessed using part B of the Trail Making Test (TMTB) while global disability was evaluated by the modified Rankin Scale (mRS).[4,9–11]

Standard protocol approvals, registrations, and patient consents. A local ethics committee validated the study in both centers. Written informed consent was obtained from the study participant or a close relative if the patient was too severely disabled.

MRI acquisition and image processing. Details of the MRI protocol are provided as supplemental data at Neurology.org. Volume of lacunes (LL_V) and that of white matter hyperintensities of presumed vascular origin (WMH_V) as well as number of microbleeds (MB_N) were measured in all participants from 3D T1, fluid-attenuated inversion recovery, and T2* sequences, respectively, using validated methods detailed previously and in agreement with the STRIVE criteria.[12,13] Brain parenchymal fraction (BPF) was defined as the ratio of brain tissue volume to the intracranial cavity volume.[12]

Statistical analyses. The aim of the present study was to answer the following question: "Given the data available today, can I predict using a simple algorithm how the clinical status of this patient will evolve over the next 3 years?" For this purpose, different linear regression models of increasing complexity and capable of predicting, at the individual level, changes in cognitive scores and disability scales (MMSE, MDRS, TMTB, and mRS) between baseline and follow-up were tested. The following variables all obtained at baseline were selected as potential predictors of clinical progression in agreement with previous reports on SVD or CADASIL[3,5]: clinical scales (MMSE, MDRS, TMTB, and mRS), epidemiologic data (age, sex, level of education), cardiovascular risk factors (smoking habits, systolic and diastolic blood pressure, alcohol consumption, levels of low-density lipoprotein cholesterol, homocysteine, HbA1C, and

C-reactive protein), and quantitative MRI markers (BPF, LL_V, WMH_V, and MB_N). A history of stroke was not included as an independent predictor as it was not associated with clinical severity in CADASIL in previous cross-sectional or longitudinal analyses.[3,5]

Because there is a risk of overfitting with linear regression, which increases with the number of predictors, the models in this study were built using Elastic Net, a penalized linear regression model algorithm, strongly limiting the risk of inclusion of a parameter not actually linked to the clinical progression.[14] The hierarchy between the different sets of variables was determined according to their expected strength of association with disease severity in CADASIL.[3,5] In model 1, only the baseline value of the considered score was included as a potential predictor (reference model; for instance, for MMSE, this model tested whether baseline MMSE was significantly predicting changes in MMSE during follow-up). In model 2, the 3 potential predictors were the baseline values of the considered score, BPF, and LL_V, the 2 MRI markers most consistently related to clinical severity in CADASIL.[4,15] In model 3, the 5 potential predictors were the baseline values of the considered score and of all quantitative MRI markers (BPF, LL_V, as well as WMH_V and MB_N, these last 2 variables being inconsistently associated with clinical severity in previous literature in CADASIL[4,10,16]). In model 4, 15 potential predictors were included: baseline values of the 4 clinical scores, of epidemiologic data, and of cardiovascular risk factors. In model 5, 17 potential predictors were included: baseline values of all predictors of model 4 together with that of BPF and of LL_V. In model 6, the 19 potential predictors were baseline values of all clinical scores, epidemiologic data, cardiovascular risk factors, and quantitative MRI markers (table 1).

Because multivariate prediction rules must be evaluated on independent datasets to avoid overly optimistic evaluation of statistical models,[17,18] we used a leave-one-out bootstrap resampling with 200 replications to estimate the accuracy of predictive models (see supplemental data for details). Two measures of prediction performances were calculated: the mean squared prediction error (MSPE) and the R^2 (coefficient of determination or ratio of explained variance) providing a non-optimistically biased assessment of our model's performances. Since R^2 values were each time estimated on independent samples including only data of patients that were not used to build the predictive models, negative values could be observed. Such values would imply that the model has a lower predictive ability than the simplest model (i.e., the mean change of the score to predict at the cohort level).

Estimates of the 2 prediction measures (MSPE, R^2) were obtained by averaging them across bootstrap replications. Standard errors and confidence intervals (CI) were obtained by calculating the SDs as well as the 2.5th and 97.5th percentiles of the bootstrapped distributions.[19] The significance of R^2 was derived from the CI obtained with bootstrapping: the null hypothesis corresponding to a mean R^2 of 0 was rejected when the CI did not include zero. Similarly, the significance of the difference in performance between 2 models was derived from the bootstrapped CI of the difference of performance measures. Finally, Bayesian Information Criterion (BIC) was used to determine the simplest model (with the lowest BIC) among those with significant but similar R^2 values.[20]

As results from linear regression models do not necessarily have a straightforward interpretation, we used recursive partitioning regression tree to determine prediction rules that can be used in daily clinical practice.[21,22] We used such algorithms to partition patients into a maximum of 3 subgroups with homogeneous changes during follow-up (will improve, remain stable, or decline)

Table 1	Baseline values of the different potential predictors[a]	
	Mean (SD) or n (%)	Min-max
Age, y	50 (10.7)	23-77.5
Sex		
Female	120 (55)	
Male	98 (45)	
Level of education[b]		
0-3	26 (12)	
4-6	144 (67)	
7-9	46 (21)	
Systolic blood pressure, mm Hg	126.8 (14.6)	94-192
Diastolic blood pressure, mm Hg	75.6 (10.0)	45-104
Smoking		
Current	45 (21)	
Never	156 (72)	
Alcohol consumption		
None	67 (31)	
<2 drinks/day	118 (54)	
>2 drinks/day	17 (8)	
LDL cholesterol, mmol/L	3.40 (0.9)	1.1-6.2
Homocysteine, mmol/L	13.17 (3.9)	5.1-27.9
HBA1C, %	5.48 (0.5)	4.4-8.4
CRP, g/L	4.78 (6.8)	1-66
MMSE	27.2 (4.0)	7-30
MDRS	136.3 (15.6)	39-144
TMTB, time to complete, s	116.6 (92.7)	24-618
mRS	0.8 (1.2)	0-5
BPF, %	86.3 (5.9)	65.3-96.5
LL_V, mm³	349.7 (625.4)	0-3,864.4
WMH_V, mm³	94,986.9 (68,258.2)	659.1-331,767.0
MB_N	1.6 (5.5)	0-51

Abbreviations: BPF = brain parenchymal fraction; CRP = C-reactive protein; LDL = low-density lipoprotein; LL_V = volume of lacunes; MB_N = number of cerebral microbleeds; MDRS = Mattis Dementia Rating Scale; MMSE = Mini-Mental State Examination; mRS = modified Rankin Scale; TMTB = Trail-Making Test, version B; WMH_V = volume of white matter hyperintensities of presumed vascular origin.

[a] Baseline predictors for the 218 patients of the Paris-Munich cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy cohort who were followed during at least 3 years.

[b] Reference levels of education: 0 = illiterate, 3 = incomplete secondary school (<9 years), 6 = secondary school (13 years), 7 = university (≥16 years).

based on the predictors' values. Briefly, this nonparametric multivariate method recursively splits the predictor space into a given number of regions containing observations with similar response values (i.e., clinical change). The algorithm determines the thresholds that identify the different groups through the maximization of an analysis of variance criterion. Additional pruning/merging was applied to the tree to obtain an intuitive rule of the clinical change.

All statistical analyses were performed with R software (https://www.r-project.org). Recursive-partitioning regression trees were built using the rpart R package.

RESULTS **Study population characteristics.** Among the 236 patients for whom a 3-year follow-up period could be achieved at the time of this study, 10 had no available quantitative MRI markers at baseline (5 did not undergo MRI, 5 had MRI of unsuitable quality for complete postprocessing) and 8 were excluded because of a large artery ischemic lesion detected on MRI at baseline, leaving 218 patients for analyses (table 1). Because of missing data in some patients, numbers of patients with full data available for analyses varied depending on the considered score (table e-1).

Predicting changes of clinical status based on linear regression models. Linear models predicting further score changes with highly significant R^2 were obtained for all outcomes (figure 1). For MDRS, TMTB, and mRS, the baseline value alone (model 1) was a significant predictor of further score changes during follow-up. By contrast, the baseline value of MMSE was not.

For MMSE, MDRS, and mRS, but not for TMTB, models including BPF and LL_V (model 2 or 5) performed significantly better than those without (models 1 and 4).

With the exception of MMSE, models based on complex sets of predictors (models 4, 5, or 6) showed lower R^2 than those based on few variables (models 2 or 3). For all 4 scores, models 2 and 3 performed very similarly, with few if any predictive value associated with the inclusion of WMH_V and MB_N.

For all 4 scores, models based on the baseline score value, BPF, and LL_V (model 2) were the models with the best tradeoff between prediction ability and number of predictors, as illustrated by their lowest BIC (table 2).

Predicting changes in clinical status based on recursive partitioning. Given the consistency of results obtained with BPF and LL_V and the fact that for all scores, model 2 obtained the lowest BIC, recursive-partitioning algorithms were built only with the baseline values of the considered score of BPF and of LL_V as predictors. This allowed identifying, according to different threshold values of the predictors, 3 groups of patients (will decline, improve, or remain stable) for MMSE, TMTB, and mRS, while only 2 groups of patients were identified for MDRS (will decline or remain stable) (figure 2). The performances of these recursive partitioning tree rules were compared to that of linear regression models (figure 1 and table e-2). Estimates of the prediction errors of these models at the individual level are shown in figure 3.

DISCUSSION We validated models predicting how the clinical status of a given patient with CADASIL

Figure 1   Ability of the different models to predict score changes over 3 years of follow-up
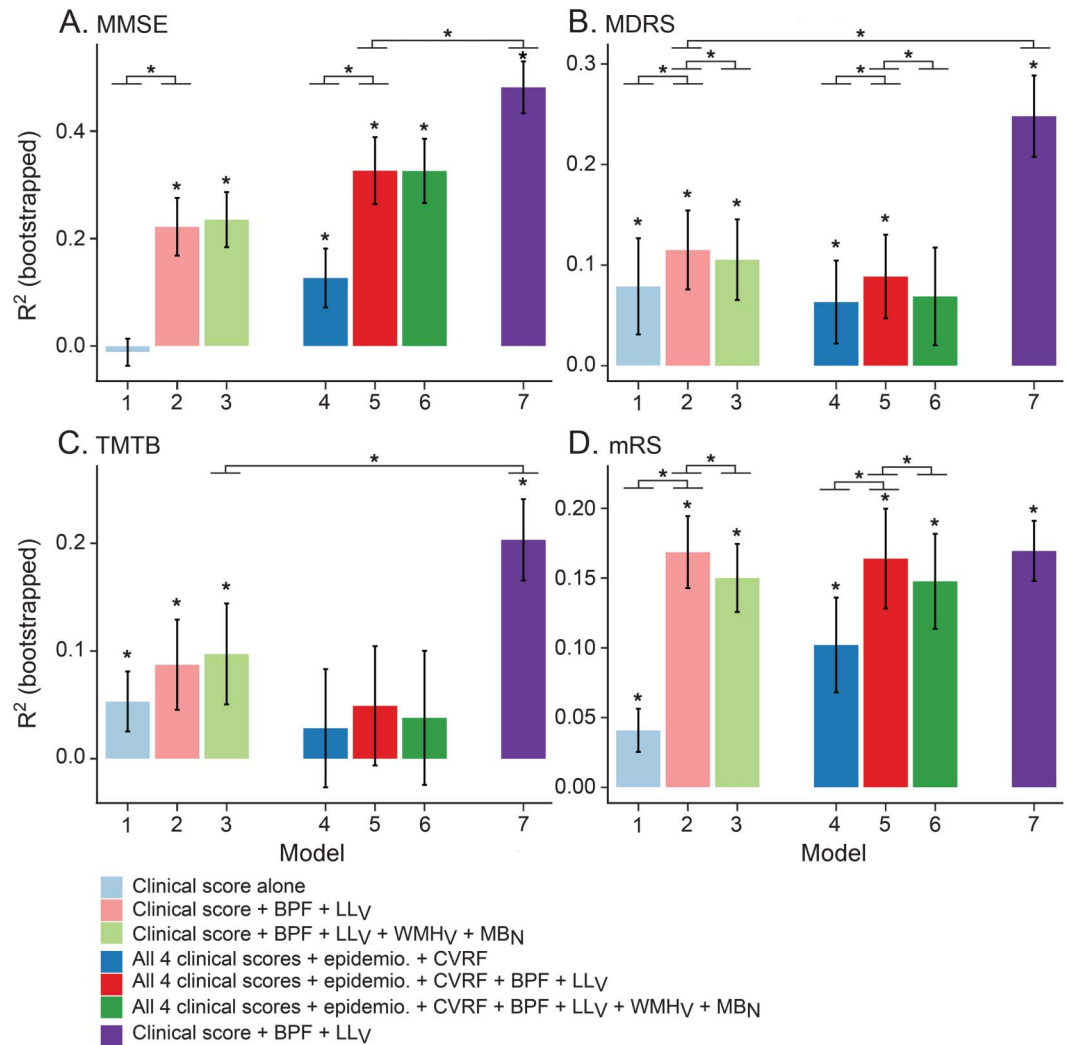
(A) Mini-Mental State Examination (MMSE), (B) Mattis Dementia Rating Scale (MDRS), (C) Trail Making Test part B (TMTB), (D) modified Rankin Scale (mRS). Estimation of the coefficient of determination (R^2) of the 7 different predictive models of score changes. Models 1 to 6 are linear models, model 7 is a recursive partitioning tree (RPT) model. *Significant R^2 differences. We performed pairwise comparisons of all linear sequentially nested models (1 vs 2, 2 vs 3, 4 vs 5, 5 vs 6) and the comparison of the RPT model vs the best linear model. BPF = brain parenchymal fraction; CVRF = cardiovascular risk factors (smoking habits, systolic and diastolic blood pressure, alcohol consumption, levels of low-density lipoprotein cholesterol, homocysteine, HbA1C, and C-reactive protein); epidemio = epidemiologic data (age, sex, level of education); LL_V = volume of lacunes; MB_N = number of cerebral microbleeds; WMH_V = volume of white matter hyperintensities of presumed vascular origin.

will evolve over the next 3 years. One of our main findings is that quantitative MRI markers (i.e., BPF and LL_V) appeared as stronger predictors of further clinical course at the individual level than any other variable including baseline clinical status, age, sex, and level of education.

The significantly lower R^2 of models including additional measures known to be associated with disease severity at the group level may appear surprising (figure 1). We could not exclude that some variables have a predictive ability in specific subsamples of patients. For instance, WMH_V may predict score changes in mildly affected patients, but not in severely

affected patients. In such cases, WMH_V may appear as a significant predictor if a randomly selected training sample includes more mildly affected participants. Then, when calculating the model R^2 on the validation samples, it would be lower because of the reduced possibility to generalize to another sample including more severe disease. In line with this hypothesis, when building models on German patients, who had generally less severe disease, WMH_V was more often retained as a significant predictor but the ability to generalize to French patients was lower (data not shown). In addition, variability and limited changes of clinical scores over a 3-year period may

Table 2 Prediction accuracy of the different models

Score	Model	Mean squared prediction error			R^2			
		Estimate	SE	95% CI	Estimate	SE	95% CI	BIC
MMSE	1	5.407	0.827	3.776–6.845	−0.012	0.025	−0.072 to 0.02	902
	2	4.224	0.56	3.126–5.335	0.222	0.054	0.12 to 0.312	847[a]
	3	4.148	0.567	3.114–5.419	0.235	0.051	0.131 to 0.323	852
	4	4.713	0.795	3.216–6.25	0.127	0.055	0.026 to 0.215	924
	5	3.627	0.51	2.706–4.736	0.327	0.062	0.203 to 0.424	866
	6	3.64	0.511	2.815–4.82	0.326	0.06	0.213 to 0.426	875
	7	2.795	0.372	2.141–3.526	0.482	0.048	0.389 to 0.555	774[b]
MDRS	1	55.06	11.05	35.76–74.31	0.079	0.048	0.018 to 0.15	1,272
	2	53.22	10.47	34.48–73.39	0.115	0.039	0.03 to 0.174	1,267[a]
	3	53.59	10.68	36.63–77.12	0.105	0.04	0.016 to 0.163	1,279
	4	55.03	11.53	34.23–78.03	0.063	0.041	0.01 to 0.129	1,335
	5	52.86	11.31	33.05–76.00	0.089	0.042	0.003 to 0.15	1,338
	6	54.06	11.41	34.66–77.83	0.069	0.049	−0.026 to 0.146	1,347
	7	44.46	8.5	30.17–61.63	0.248	0.041	0.164 to 0.316	1,246[b]
TMTB	1	3,127.7	663.0	2,023.9–4,298.4	0.053	0.028	0.006 to 0.1	1,908
	2	3,069.4	593.3	2,069.3–4,252.2	0.087	0.042	0.002 to 0.148	1,901[a]
	3	2,977.5	618.9	1,983.5–4,336.7	0.097	0.047	0.001 to 0.163	1,904
	4	3,198.6	689.7	2,022.7–4,534.5	0.028	0.055	−0.076 to 0.113	1,962
	5	3,186.1	659.8	1,972.1–4,421.9	0.049	0.055	−0.051 to 0.145	1,962
	6	3,216.5	710.4	1,978.7–4,605.9	0.038	0.062	−0.078 to 0.137	1,970
	7	2,707.5	509.4	1,785.4–3,845.4	0.203	0.038	0.12 to 0.258	1,885[b]
mRS	1	0.65	0.079	0.507–0.787	0.041	0.015	0.007 to 0.063	502
	2	0.565	0.069	0.432–0.691	0.169	0.026	0.114 to 0.211	477[a,b]
	3	0.578	0.072	0.437–0.706	0.15	0.024	0.105 to 0.192	487
	4	0.608	0.079	0.458–0.744	0.102	0.034	0.045 to 0.157	537
	5	0.569	0.073	0.441–0.704	0.164	0.036	0.098 to 0.222	522
	6	0.582	0.071	0.454–0.703	0.148	0.034	0.085 to 0.204	537
	7	0.565	0.069	0.442–0.688	0.169	0.022	0.128 to 0.204	481

Abbreviations: BIC = Bayesian information criterion; CI = confidence interval; MDRS = Mattis Dementia Rating Scale; MMSE = Mini-Mental State Examination; mRS = modified Rankin Scale; R^2 = coefficient of determination; TMTB = Trail Making Test part B.
Prediction accuracy of the 7 different models evaluated using mean squared prediction error, R^2 and Bayesian information criterion. Models 1 to 6 refer to linear regression models based on different sets of predictors; model 7 refers to the recursive partitioning tree models (figure 1).
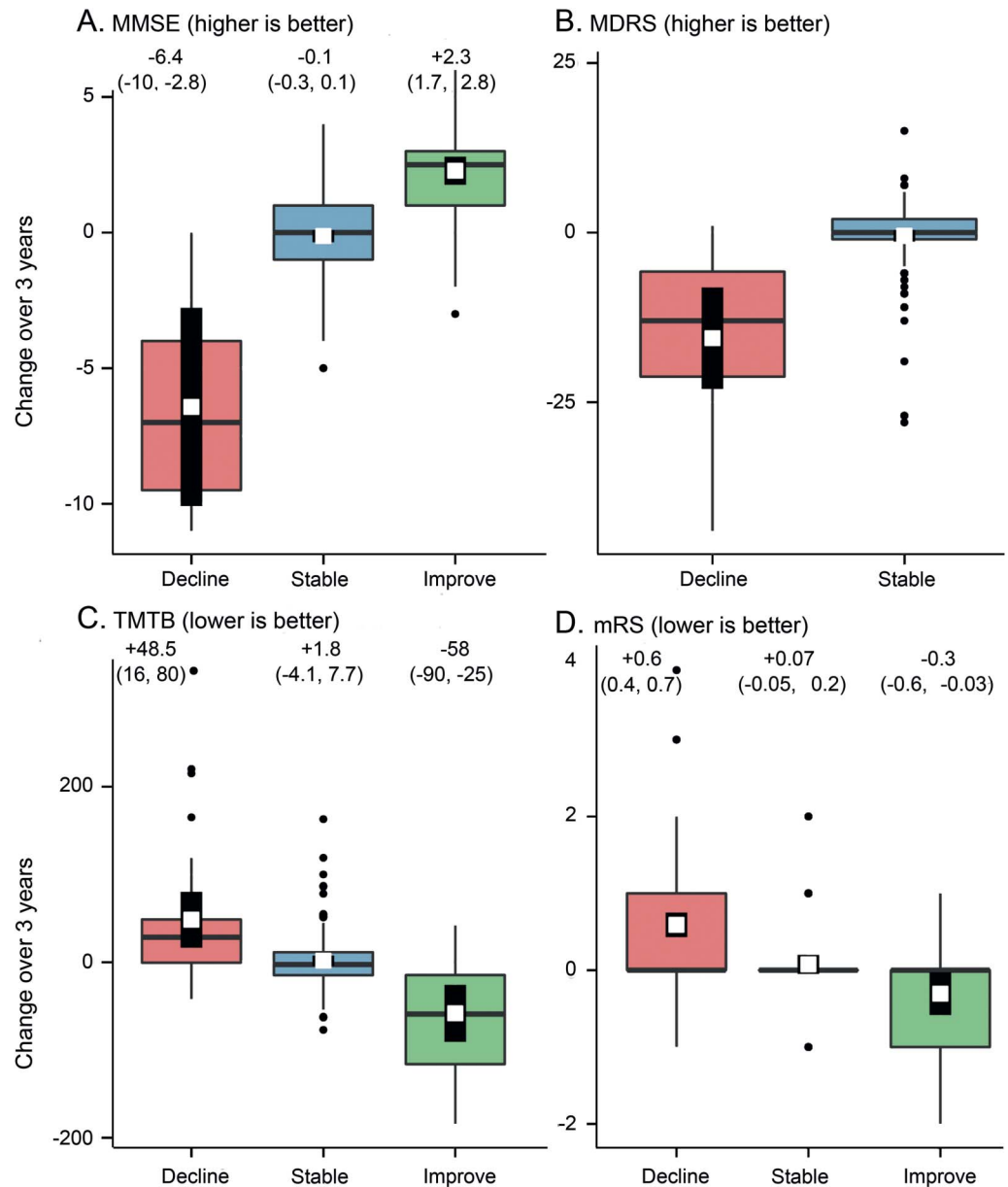[a] The best linear regression model for each considered score.
[b] The best model for each considered score (either linear regression model or recursive partitioning tree model).

prevent the detection of small effects related to multiple factors. Alternatively, LL_V and BPF may substitute for the combined and long-term consequences of age, sex, or cardiovascular risk factors on the cerebral tissue.[3,5]

Given the low amplitude of score changes during follow-up compared to score ranges, the variance explained by models predicting follow-up score values would have been artificially high. When facing 2 patients with an MMSE score of 26, demonstrating that their follow-up score will lie between 23 and 29 with a high probability is of limited interest. By contrast,

determining that 1 patient is at high risk to lose 6 points (thus scoring 20) and the other to increase its score by 2 (thus scoring 28) appears of higher interest. Partitioning-tree algorithms both confirmed the results obtained with linear regression models and allowed groups of patients with similar prediction of further evolution to be identified. For instance, we found that a patient having an LL_V above 1,964 mm³ will lose 6 points on MMSE, regardless of any other variable known to be associated with the disease severity. By contrast, a patient with an LL_V inferior to 1,964 mm³, furthermore in the absence of
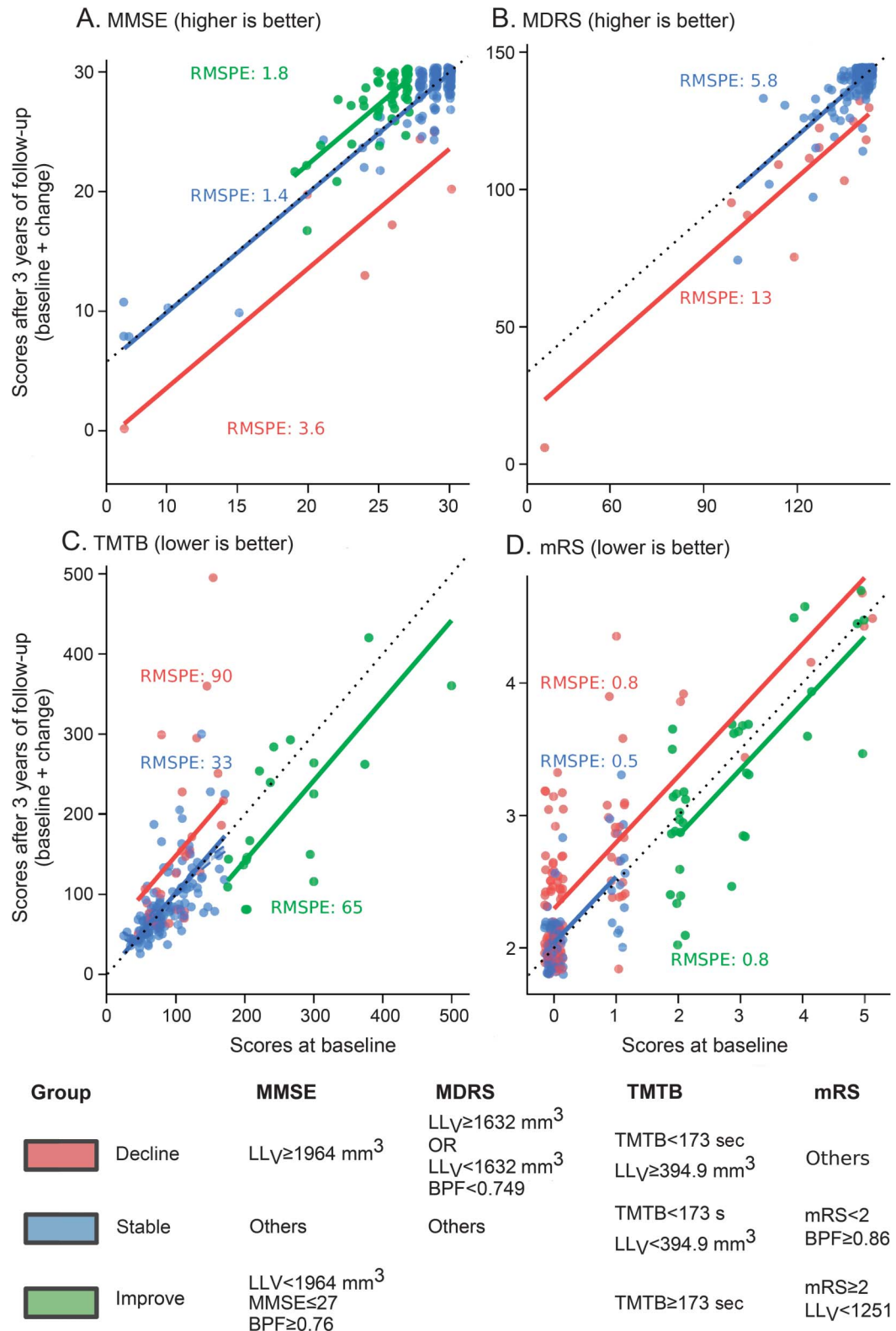
Figure 2   Simple rules for predicting clinical changes

A. MMSE (higher is better)

-6.4	-0.1	+2.3
(-10, -2.8)	(-0.3, 0.1)	(1.7, 2.8)

Change over 3 years

Decline   Stable   Improve

B. MDRS (higher is better)

Decline   Stable

C. TMTB (lower is better)

+48.5	+1.8	-58
(16, 80)	(-4.1, 7.7)	(-90, -25)

Change over 3 years

Decline   Stable   Improve

D. mRS (lower is better)

+0.6	+0.07	-0.3
(0.4, 0.7)	(-0.05, 0.2)	(-0.6, -0.03)

Decline   Stable   Improve

Group		MMSE	MDRS	TMTB	mRS
	Decline	$LL_V \geq 1964$ mm^3	$LL_V \geq 1632$ mm^3 OR $LL_V < 1632$ mm^3 BPF<0.749	TMTB<173 sec $LL_V \geq 394.9$ mm^3	Others
	Stable	Others	Others	TMTB<173 s $LL_V < 394.9$ mm^3	mRS<2 BPF≥0.86
	Improve	LLV<1964 mm^3 MMSE≤27 BPF≥0.76		TMTB≥173 sec	mRS≥2 LL$_V$<1251

(A) Mini-Mental State Examination (MMSE), (B) Mattis Dementia Rating Scale (MDRS), (C) Trail Making Test part B (TMTB), (D) modified Rankin Scale (mRS). Box plots of score changes between baseline and follow-up. The horizontal band inside the box is the median. The overlaid white square and the thick vertical line depict the prediction and its 95% confidence interval made by recursive partitioning tree (RPT) model at the group level. The RPT predictive model has partitioned patients into homogeneous subgroups according to the baseline values of brain parenchymal fraction (BPF), volume of lacunes (LL$_V$), and the considered score at baseline. Clinical evolution as estimated by the model is color-coded (pink: decline; blue: stable; green: improve). Numbers indicate the average change within the considered group with 95% confidence intervals.

Figure 3 Follow-up scores as a function of baseline scores according to patient group

Group		MMSE	MDRS	TMTB	mRS
🟥	Decline	LL$_V$≥1964 mm^3	LL$_V$≥1632 mm^3 OR LL$_V$<1632 mm^3 BPF<0.749	TMTB<173 sec LL$_V$≥394.9 mm^3	Others
🟦	Stable	Others	Others	TMTB<173 s LL$_V$<394.9 mm^3	mRS<2 BPF≥0.86
🟩	Improve	LLV<1964 mm^3 MMSE≤27 BPF≥0.76		TMTB≥173 sec	mRS≥2 LL$_V$<1251

(A) Mini-Mental State Examination (MMSE), (B) Mattis Dementia Rating Scale (MDRS), (C) Trail Making Test part B (TMTB), (D) modified Rankin Scale (mRS). Three-year follow-up scores as a function of the baseline value. Each single point corresponds to 1 patient, on the horizontal axis the initial value of the considered score and on the vertical axis the corresponding value after 3 years of follow-up (a small amount of jitter has been introduced to help read similar values, explaining why some points can appear to have decimal values). Solid color lines correspond to predictions obtained by recursive partitioning tree models. The predicted value of the considered score after 3 years is obtained by projecting onto the vertical axis

Continued

cerebral atrophy, will either keep a score of 26 or even improve up to 2 points.

Our study has several limitations. Since the models were based on patients with both baseline and follow-up data, our algorithms may not be applicable to every CADASIL case but only to patients who were not at the most advanced stage of the disease and who came back to a distant clinical center for their follow-up evaluation. Among the 54 patients who could not achieve a 3-year follow-up period, 23 were severely disabled or had dementia, 14 died, but 17 of them did not attend for other reasons, including unwillingness to pursue the follow-up. Another limitation is that our models predict clinical scores rather than clinical status. Variations related to the scoring method itself such as learning or ceiling effects clearly appear in our results. A patient with an MMSE score of 28 that can increase by 2 points is compared to another patient whose score is 30 that cannot increase. Specific processes may also intervene differently for the various scores. For instance, for TMTB, the first order thresholds for identifying different groups in the recursive-partitioning approach was the baseline value of the score itself, with the best improvement in patients achieving the test in more than 173 seconds. These results are probably secondary to the test itself rather than to clinical evolution, as there is more space for these patients to improve their performances in case they did not perform well at the first visit. TMTB is probably the harder score to predict, as shown by its lower R^2 compared to other scores. This is in line with the possible variations from one visit to another on TMTB, depending on many factors not necessarily linked to clinical severity, including fatigue and mood.[23] We do not see any reason why TMTB would not be altered from learning effects, but it seems that the variability in patient performances on this test outweighs this effect.

In the next few years, one can imagine that after automated postprocessing, the obtained values of BPF and LL_V would be fed to simple software providing in return a clinical prediction far more precise than that achieved by eye. Meanwhile, the algorithm could still be used through visual estimation of these values (figures e-1 and e-2 illustrate the values of BPF and of LL_V corresponding to the thresholds obtained in the recursive-partitioning algorithm).

We did not include in the present models more elaborated methods to take into account the effect of white matter hyperintensities in specific locations,[10] the occurrence of secondary degeneration, or the occurrence of cortical alterations.[24–26] We made this choice to ensure the best generalizability of these models and keep the analyses reasonably simple and translatable in clinical practice. Another limitation is the different sizes of the samples available to predict the 4 scores, which may have influenced our results. Finally, we chose to constrain the behavior of the recursive partitioning tools to identify a maximum of 3 groups based on a maximum of 3 different baseline variables. It is important to keep in mind that, as in every nonlinear model, recursive partitioning trees are prone to overfitting, which risks translating into loss of generalizability. Here, we obtained strong arguments with our validation scheme to prove the validity of our approach.

This study has several strengths. The results were obtained using a large sample of patients with CADASIL who were followed in 2 different referral centers in 2 different countries. Thus, they are likely to be generalized to new populations of patients with CADASIL in similar countries. Our specific statistical approach obtained validated models that were compared using state-of-the-art methodology. The method we chose also prevented overfitting the data and provided a true demonstration of the ability of these models to generalize to new patients with CADASIL. Our specific strategy identified markers not only associated with the disease evolution at the group level, but true predictors of further disease evolution at the individual level. It also demonstrated that, in order to predict the disease evolution for a given patient, the interest of clinical variables appears limited and most information may be contained in quantitative MRI markers that integrate both aging and disease effects.

AUTHOR CONTRIBUTIONS

Eric Jouvent: data acquisition, study design, statistical analyses, drafting manuscript. Edouard Duchesnay: study design, statistical analyses. Foued Hadj-Selem: statistical analyses. François De Guio: study design, data acquisition. Jean-François Mangin: study design. Dominique Hervé: data acquisition. Marco Duering: study design, data acquisition. Stefan Ropele: study design, manuscript correction. Reinhold Schmidt: study design, manuscript correction. Martin Dichgans: study design, data acquisition, data analysis. Hugues Chabriat: study design, data acquisition, data analysis, manuscript correction.

Figure 3 legend, continued:
the intersection between the baseline value on the horizontal axis and the solid line corresponding to the patient's group. The black dashed line stands for stability, where follow-up score equals baseline. For each group, we report the root mean squared prediction error (RMSPE), which is an estimate of the prediction error calculated from 200 bootstrapped replications. As illustrated by the dispersion of colored points around solid lines, the prediction error at the individual level is inversely related to the sample size. LL_V = volume of lacunes.

REFERENCES

1. Joutel A, Corpechot C, Ducros A, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 1996;383:707–710.
2. Zieren N, Duering M, Peters N, et al. Education modifies the relation of vascular pathology to cognitive function: cognitive reserve in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. Neurobiol Aging 2013;34:400–407.
3. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Bousser MG. CADASIL. Lancet Neurol 2009;8: 643–653.
4. Viswanathan A, Godin O, Jouvent E, et al. Impact of MRI markers in subcortical vascular dementia: a multimodal analysis in CADASIL. Neurobiol Aging 2010;31: 1629–1636.
5. Chabriat H, Herve D, Duering M, et al. Predictors of clinical worsening in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: prospective cohort study. Stroke 2016;47:4–11.
6. Viswanathan A, Gschwendtner A, Guichard JP, et al. Lacunar lesions are independently associated with disability and cognitive impairment in CADASIL. Neurology 2007;69:172–179.
7. Benisty S, Reyes S, Godin O, et al. White-matter lesions without lacunar infarcts in CADASIL. J Alzheimers Dis 2012;29:903–911.
8. Jouvent E, Reyes S, Mangin JF, et al. Apathy is related to cortex morphology in CADASIL A sulcal-based morphometry study. Neurology 2011;76:1472–1477.
9. Dichgans M, Markus HS, Salloway S, et al. Donepezil in patients with subcortical vascular cognitive impairment: a randomised double-blind trial in CADASIL. Lancet Neurol 2008;7:310–318.
10. Duering M, Gonik M, Malik R, et al. Identification of a strategic brain network underlying processing speed deficits in vascular cognitive impairment. Neuroimage 2013; 66:177–183.
11. Epelbaum S, Benisty S, Reyes S, et al. Verbal memory impairment in subcortical ischemic vascular disease: a descriptive analysis in CADASIL. Neurobiol Aging 2011;32:2172–2182.
12. Jouvent E, Viswanathan A, Mangin JF, et al. Brain atrophy is related to lacunar lesions and tissue microstructural changes in CADASIL. Stroke 2007;38:1786–1790.
13. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol 2013;12:822–838.
14. Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 2005;67: 301–320.
15. Liem MK, van der Grond J, Haan J, et al. Lacunar infarcts are the main correlate with cognitive dysfunction in CADASIL. Stroke 2007;38:923–928.
16. Liem MK, Lesnik Oberstein SAJ, Haan J, et al. MRI correlates of cognitive decline in CADASIL: a 7-year follow-up study. Neurology 2009;72:143–148.
17. Vapnik V. The Nature of Statistical Learning Theory. Berlin: Springer; 2000.
18. Ambroise C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. Proc Nat Acad Sci 2002;99:6562–6566.
19. Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1:54–75.
20. Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–464.
21. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. Boca Raton, FL: CRC Press; 1984.
22. Fonarow GC, Adams KF Jr, Abraham WT, et al. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA 2005;293:572–580.
23. Jouvent E, Reyes S, De Guio F, Chabriat H. Reaction time is a marker of early cognitive and behavioral alterations in pure cerebral small vessel disease. J Alzheimers Dis 2015;47:413–419.
24. Duering M, Righart R, Csanadi E, et al. Incident subcortical infarcts induce focal thinning in connected cortical regions. Neurology 2012;79:2025–2028.
25. Jouvent E, Poupon C, Gray F, et al. Intracortical infarcts in small vessel disease a combined 7-T postmortem MRI and neuropathological case study in cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy. Stroke 2011;42:E27–E30.
26. De Guio F, Reyes S, Vignaud A, et al. In vivo high-resolution 7 Tesla MRI shows early and diffuse cortical alterations in CADASIL. PLoS One 2014;9:e106311.