AJNR

Final Cerebral Infarct Volume Is Predictable by MR Imaging at 1 Week

T. Tourdias, P. Renou, I. Sibon, J. Asselineau, L. Bracoud, M. Dumoulin, F. Rouanet, J.M. Orgogozo and V. Dousset

T. Tourdias
P. Renou
I. Sibon
J. Asselineau
L. Bracoud
M. Dumoulin
F. Rouanet
J.M. Orgogozo
V. Dousset

Final Cerebral Infarct Volume Is Predictable by MR Imaging at 1 Week

BACKGROUND AND PURPOSE: Stroke volume, an increasingly used end point in phase II trials, is considered stationary at least 30 days after the ictus. We investigated whether information conveyed by MR imaging measurements of the "final" infarct volume could be assessed as early as the subacute stage (days 3–6), rather than waiting for the chronic stage (days 30–45).

MATERIALS AND METHODS: Ninety-five patients with middle cerebral artery stroke prospectively included in a multicenter study underwent MR imaging during the first 12 hours (MR imaging-1), between days 3 and 6 (MR imaging-2), and between days 30 and 45 (MR imaging-3). We first investigated the relationship between subacute (FLAIR-2) and chronic volumes (FLAIR-3), by using a linear regression model. We then tested the relationship between FLAIR volumes (either FLAIR-2 or FLAIR-3) and functional disability, measured by the mRS at the time of MR imaging-3, by using logistic regression. The performances of the models were assessed by using the AUC in ROC.

RESULTS: A linear association between log FLAIR-2 and log FLAIR-3 volumes was observed. The proportion of FLAIR-3 variation, explained by FLAIR-2, was high ($R^2 = 81\%$), without a covariate that improved this percentage. Both FLAIR-2 and FLAIR-3 were independent predictors of mRS (OR, 0.79 and 0.73; 95% CI, 0.64–0.97 and 0.56–0.96; $P = .026$ and .023). The performances of the models for the association between either FLAIR volume and mRS did not differ (AUC = 0.897 for FLAIR-2 and 0.888 for FLAIR-3).

CONCLUSIONS: Stroke damage may be assessed by a subacute volume because subacute volume predicts the "true" final volume and provides the same clinical prognosis.

ABBREVIATIONS: ADC = apparent diffusion coefficient; AUC = area under the curve; BI = Barthel index; CI = confidence interval; DWI = diffusion-weighted imaging; ECASS = European Cooperative Acute Stroke Study; EPITHET = Echo-Planar Imaging Thrombolytic Evaluation Trial; FLAIR = fluid-attenuated inversion recovery; ICA = internal carotid artery; log = logarithmic; MRA = MR angiography; mRS = modified Rankin Scale, NIHSS = National Institutes of Health Stroke Scale; OR = odds ratio; PH = parenchymal hematoma; PWI = perfusion-weighted imaging; Q1-Q3 = first and third quartile of interquartile range; ROC = receiver operating characteristic analysis; TP = time point; TTP = time-to-peak; VIRAGE = Valeur predictive des paramètres IRM à la phase aigue de l'Accident vasculaire cerebral: application à la Gestion des Essais thérapeutiques

Infarct volume is a direct measurement of 1 of the final pathologic steps leading to the clinical deficits caused by an ischemic stroke.[1] Therefore, final infarct volume derived from MR imaging represents an objective and valid measurement of stroke consequences. Infarct volume has been proposed as an alternative (surrogate end point) to classic disability or handicap scales,[2,3] being able to improve the efficacy of phase II trials by increasing their statistical power and, therefore, their

ability to reach conclusions about the potential benefits of therapy.[4] The same marker is also frequently used in conceptual studies, particularly in penumbra proof-of-concept studies, which consider the fate of "at risk" tissue as an outcome variable.[5]

Nevertheless, there is currently no consensus as to the appropriate timing for final infarct volume measurements. Typically, outcome scans are performed relatively late, between 30 days[6] and 90 days[2,7] following the insult when the infarct volume is considered as fixed. It was recently shown that changes in volume between day 30 and day 90 were insignificant, proving that a time span of 30 days following ictus is sufficient to assess final infarct volume.[8] An earlier measurement, conducted during the first week, is thought to overestimate stroke lesions due to vasogenic edema.[9] Nevertheless, such early measurements, taken before the patient's discharge from the hospital, could greatly facilitate future studies and trials because this early end point could decrease the study duration, costs, and the number of patients lost to follow-up.

Our study's aim was to evaluate whether information provided by MR imaging measurements of "final" cerebral infarct volume could be assessed as early as the subacute stage (days 3–6), rather than waiting for the chronic stage (days 30–45), thereby leading to an earlier outcome marker.

Materials and Methods

Patients

A total of 95 patients with first middle cerebral artery acute ischemic stroke were included in a prospective national multicenter observational cohort study called VIRAGE. The study was approved by the institutional review board, and written informed consent was obtained from all participants. Four university hospitals in France recruited patients between September 2003 and June 2008. Primary inclusion criteria were as follows: men and women, older than 18 years, with a clinical diagnosis of minor-to-severe cerebral infarct (NIHSS scores between 4 and 20) in the left or right middle cerebral artery territory evolving for <12 hours. Exclusion criteria were the following: coma, transient ischemic attacks or lacunar syndrome, pregnant or breast-feeding women or women without a negative pregnancy test, and contraindications to MR imaging.

Patients were examined at 3 predefined TPs following the insult: TP1, < 12 hours; TP2, between day 3 and day 6; and TP3, between day 30 and day 45. MR imaging was performed at each TP. The NIHSS Score was recorded at TP1, whereas BI and mRS were assessed at TP2 and TP3. NIHSS, BI, and mRS were evaluated by neurologists with expertise in the administration of these scales, who were unaware of the volume measurements. Clinical and biologic data were collected via electronic case report forms.

MR Imaging Protocol

MR imaging studies were performed on 1.5T magnets (Intera and Achieva, Philips Healthcare, Best, the Netherlands; and Magnetom Vision, Siemens, Erlangen, Germany). Sequence parameters were standardized across centers and summarized in detail elsewhere.[10] TP1 and TP2 scans included DWI (DWI-1 and DWI-2), dynamic susceptibility PWI (PWI-1 and PWI-2), time-of-flight MRA (MRA-1 and MRA-2), and FLAIR (FLAIR-1 and FLAIR-2). At chronic follow-up (TP3), FLAIR-3, and DWI-3 were analyzed. T2* gradient recalled-echo was also available at each TP.

Data Processing and Image Analysis

MR imaging studies were centralized by using an independent core laboratory (Bio-Imaging, Lyon, France), which developed dedicated software tools for randomized image-review sessions. All MR images were read at the coordinating center by investigators blinded to the clinical information. Results were automatically extracted by Bio-Imaging and then sent directly to the core statistical unit to ensure safety against post hoc changes because the data could no longer be modified after transfer.

DWI and FLAIR Volumes. Stroke volumes were assessed by a neuroradiologist (T.T.) outlining FLAIR and DWI abnormalities at each TP. Both sequences were viewed simultaneously on a single screen, together with ADC-1 at TP1 and ADC-2 at TP2. This allowed the reader to edit lesion borders on DWI hyperintensity with decreased ADC and FLAIR hyperintensity in the corresponding topography. For TP3, FLAIR-3 and DWI-3 were shown with FLAIR-1 and DWI-1 to edit lesion borders on FLAIR-3 hyperintensity corresponding to the infarcted territory on DWI-1, with particular attention being paid to avoid replication of pre-existing chronic lesions present on FLAIR-1. If present, the volume of hemorrhagic transformation was included, and its subtype was noted according to the ECASS radiologic classification based on a T2* gradient-recalled sequence.[11] The order of appearance on the screens was randomized so that volumes at TP2 were analyzed by readers blinded to the measurements at

TP3. All volumes were pre-detected by using an automatic unsupervised 3D segmentation algorithm (Bio-Imaging), followed by a manual validation by the reader, by using interactive drawing tools. We previously reported a very good reproducibility (intraclass correlation coefficients between 0.86 and 0.96) for such quantitative volume measurements on the first 30 patients from the VIRAGE data base.[12]

PWI Analyses. Perfusion deficit volumes were estimated by the same reader at TP1 and TP2, according to the time-to-peak maps computed by Bio-Imaging randomly displayed along with the corresponding DWI sequences. A 4-second threshold was chosen as a reliable estimate of penumbral flow,[13] according to a previously described method.[14]

MRA Analyses. A second reader (P.R.) reviewed all MRAs at TP1 and TP2 in a randomized order, including the native images reconstructed in 3 orthogonal planes, maximum intensity projections, and DWI-1 to indicate the infarct side. Scoring of the distal ICA and the middle cerebral artery M1 and M2 status was categorized as follows: 1 = normal, 2 = disturbed flow, 3 = occluded, and 4 = technically inadequate.

Statistical Analyses

The descriptive statistics used were mean ± SD or medians and quartiles (Q1–Q3). Volume comparisons were performed by using the nonparametric paired-sample Wilcoxon signed rank test.

To address the main issue, we first evaluated whether a subacute measure could predict the fixed (eg, chronic) volume and, second, whether a subacute measure could provide the same prognostic information on the "true" clinical outcome as a chronic measure could.

The association between the chronic volume (FLAIR-3) and the subacute volume (FLAIR-2) was first estimated by using a Spearman rank correlation coefficient. Then, for the prediction of FLAIR-3 by the measurement of FLAIR-2, a multivariable linear model was used according to a backward stepwise method,[15] with the independent variables listed in Table 1. Assumptions of normality and homoscedasticity of residuals (the difference between the observed values and the values predicted by the model) were graphically checked. The adjusted coefficient of determination,

---

Table 1: Independent variables[a]

Variables	
Clinical	
Age	Quantitative variable
Sex	Categorized: 1) male, 2) female
Initial glycemia	Quantitative variable
Fibrinolytic treatment	Categorized: 1) yes, 2) no
Initial NIHSS score	Quantitative variable
Imaging	
FLAIR volume	Quantitative variable
Persistent occlusion[b]	Categorized: 1) yes, 2) no
Percentage of reperfusion[c]	Categorized: 0) no significant abnormality on PWI-1,[d] 1) reperfusion between TP1 and TP2 ≥90%; 2) Reperfusion between TP1 and TP2 < 90%

[a] Independent variables were introduced in multivariable analyses because of their potential association with final infarct volume and clinical recovery.[34–36] Quantitative variables were introduced as continuous variables or by class according to quartiles if the hypothesis of linearity was not graphically met.
[b] Persistent occlusion was defined as an occluded artery on MRA-1 remaining occluded on MRA-2 in an acute stroke territory.
[c] Reperfusion definition was based on PWI at TP1 and TP2, calculated as follows: (PWI-1 − PWI-2)/PWI-1 × 100. Ninety percent was chosen as a threshold because it was previously reported to be associated with final volume and clinical outcome.[36]
[d] PWI abnormalities were considered on TTP maps using a delay of 4 seconds or more. TP1 indicates <12 hours; TP2, days 3–6.

Table 2: Patient characteristics, VIRAGE study

Characteristics	Values
Mean age (yr) (SD)	63.0 (13.1)
Sex, female (No.) (%)	37 (40.7)
Mean glycemia (mmol/L) ($\pm$SD)	7.0 (2.5)
Intravenous fibrinolytic treatment (No.) (%)	42 (46.2)
Median NIHSS score at presentation (Q1–Q3)	11 (6–17)
mRS score at TP3 (Q1–Q3)	2 (1–3)
mRS >1, (No.) (%)/mRS ≤1 (No.) (%)	54 (59.3)/36 (39.6)
mRS >2, (No.) (%)/mRS ≤2 (No.) (%)	32 (35.2)/58 (63.7)

FLAIR-1   **DWI-1**

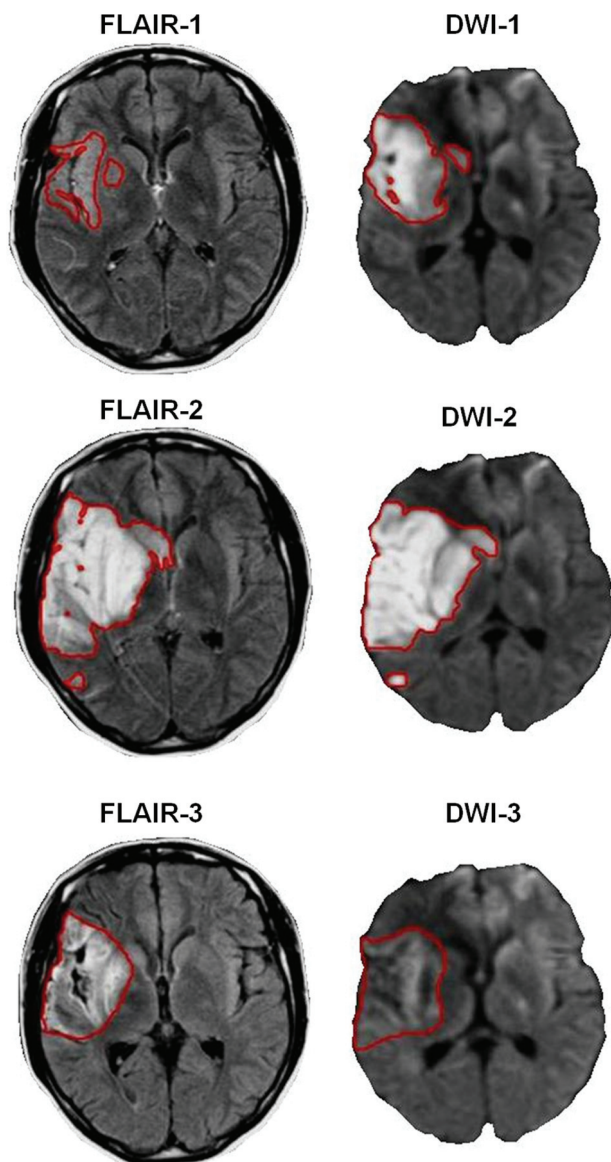**FLAIR-2**   **DWI-2**

FLAIR-3   **DWI-3**

Fig 1. Example of time course of volumes. Axial FLAIR and axial DWI acquired within the first 12 hours following stroke onset (FLAIR-1 and DWI-1), between 3 and 6 days (FLAIR-2 and DWI-2), and between 30 and 45 days (FLAIR-3 and DWI-3). Typically, the pattern of volume change includes an increase at the subacute stage and a secondary decrease between the subacute and chronic stages.

R^2, was used to assess the percentage of variability explained by the variables included in the model.

Last, the association between functional disability at days 30–45 and either FLAIR-2 or FLAIR-3 volumes was estimated by using multivariable logistic regression according to a backward stepwise meth-

Table 3: Imaging characteristics, VIRAGE study

Characteristics	Values	
	Median (Q1–Q3)	Range
DWI-1 volume (cm^3)	13.4 (4.1–34.9)	0–164.7
PWI-1 volume (cm^3)	34.9 (6.9–87.8)	0–243.8
FLAIR-2 volume (cm^3)	25 (8–60)	0–308
FLAIR-3 volume (cm^3)	20 (6–45)	0–211
Occlusion		
No persistent (No.) (%)	76 (83.5)	
Persistent (No.) (%)[a]	15 (16.5)	
ICA	9 (9.9)	
M1	4 (4.4)	
M2	2 (2.2)	
Reperfusion		
No significant perfusion abnormality at TP1, n (%)	21 (23.1)	
Reperfusion ≥90%	35 (38.5)	
Reperfusion <90%	33 (36.3)	
Technically inadequate	2 (2.2)	

[a] Persistent occlusion was defined as an occluded artery on MRA-1 remaining occluded on MRA-2 in an acute stroke territory.

od,[15] with the independent variables listed in Table 1. For this purpose, disability was assessed by mRS dichotomized at a cutoff of 1 for 2 major reasons: First, mRS was a better tool after mild stroke (as seen in the VIRAGE population) than BI[16]; and second, a full recovery defined as mRS ≤ 1 was the criterion used in the main studies demonstrating the efficacy of alteplase.[17,18] Results with mRS dichotomized at a cutoff of 2 were also presented as secondary analyses. The performances of the models were assessed by using the AUC in an ROC for both models.

All analyses were performed by using SAS Software, Version 9.1.3 (SAS Institute, Cary, North Carolina).

Results

Patient Characteristics and Neuroimaging Variables

Of the 95 patients, 4 were excluded from analyses because both MR imaging-2 and MR imaging-3 data were missing. Two patients were transferred to another hospital and lost to follow-up, 1 died, and 1 underwent pacemaker implantation after MR imaging-1, contraindicating other MR imaging scans. The remaining 91 patients underwent a total of 267 MR imaging examinations (91 MR imaging-1, 90 MR imaging-2, and 86 MR imaging-3). One MR imaging at TP2 was technically inadequate for FLAIR volume, due to motion artifacts. At TP3, 2 MR images were technically inadequate, 2 patients died before TP3, and 1 patient was not able to undergo MR imaging-3, due to clinical deterioration.

Table 2 features the patient characteristics. Neuroimaging variables were as follows: MR imaging-1 was performed after a mean delay of 6.3 ± 3.3 hours following stroke onset ($n = 91$). In total, 17 patients (18.7%) were examined during the first 3 hours, 29 (31.9%) between 3 and 6 hours, and 44 (48.4%) between 6 and 12 hours. MR imaging-2 was performed after a mean delay of 4.2 ± 1.1 days ($n = 90$), and MR imaging-3, after a mean delay of 37.9 ± 13.5 days ($n = 86$). Median lesion volume significantly increased from DWI-1 (13.4 cm^3) to FLAIR-2 (25 cm^3, $P < .05$; $n = 90$) and then significantly regressed to FLAIR-3 (20 cm^3, $P < .05$; $n = 86$). Figure 1 shows an illustrative case of the volume time course, and Table 3 provides details on volume measurements and reperfusion
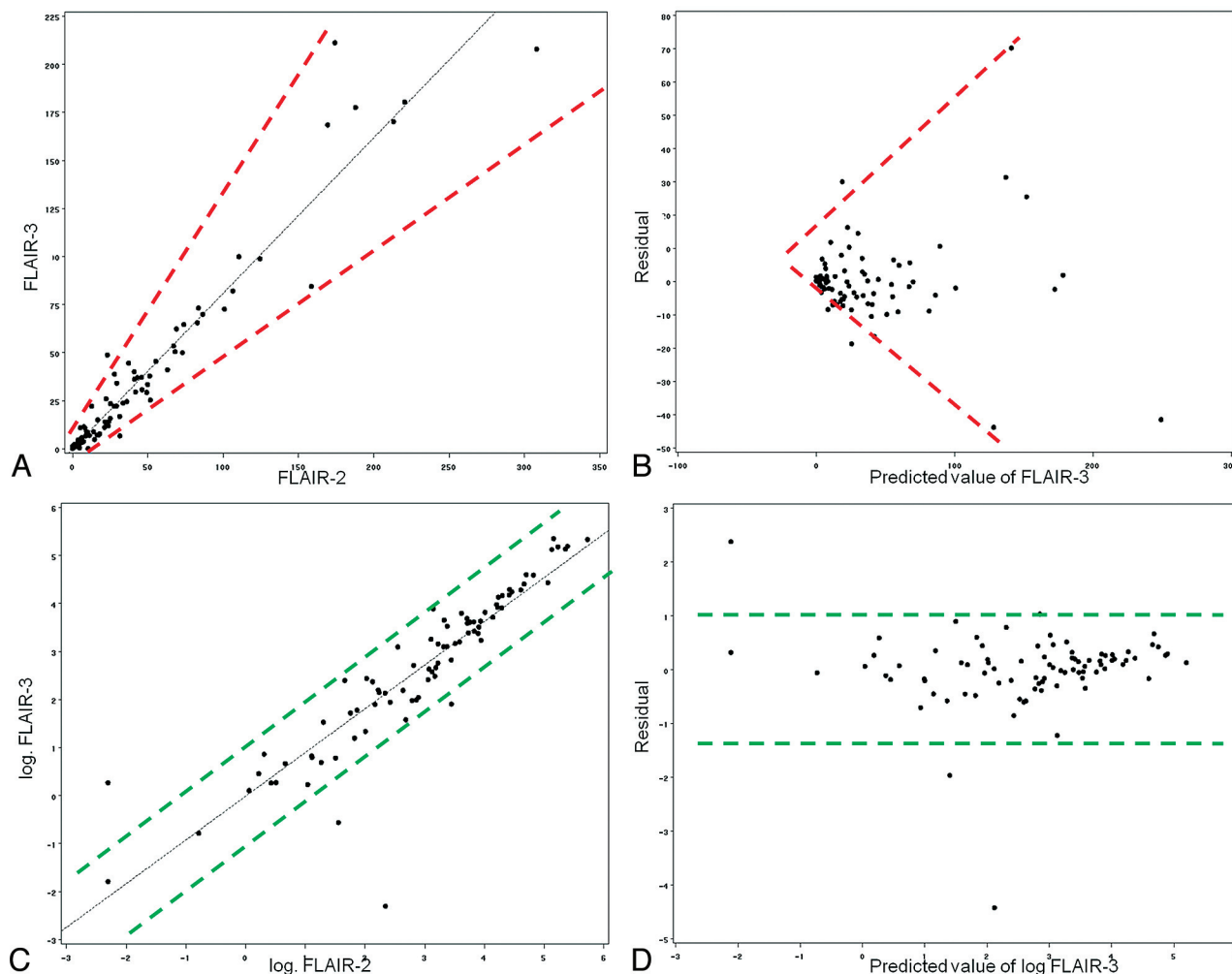
Fig 2. Scatterplots of lesion volumes. *A* and *C,* The association of FLAIR-2/FLAIR-3 and log FLAIR-2/log FLAIR-3, respectively. *B* and *D,* The representations of the residuals as a function of FLAIR-3 or log FLAIR-3 predicted values. Residuals represent the difference between the observed values (FLAIR-3 in *B* and log FLAIR-3 in *D*) and the values predicted by the model. The prediction of the model is good if residuals are concentrated around zero for the entire set of predicted values. In the case of absolute values, the more the FLAIR-2 volume increases, the less efficient is the prediction of FLAIR-3 is (*A, red broken lines*), resulting in a residual dispersion for higher volume (>100 cm^3) prediction (*B, red broken lines*: heteroscedasticity). After logarithmic transformation, there is no longer any dispersion for large volumes (*C, green broken lines*) and residuals are highly concentrated around the zero point without dispersion (*D, broken dot lines*: homoscedasticity).

data. At TP2, 30 patients exhibited hemorrhagic transformation (33%). Most had hemorrhagic petechia (HI1, $n = 15$, and HI2, $n = 8$, according to the ECASS classification), and 7 had PH (PH1, $n = 7$, and PH2, $n = 0$).

Relationship between Subacute and Chronic Volumes

The Spearman rank correlation coefficient assessing the relationship between lesion volumes at TP2 and TP3 was high (0.95, $P < .001$, $n = 85$) between FLAIR-2 and FLAIR-3. Nevertheless, for large volumes (>100 cm^3), FLAIR-2 volumes could evolve toward very different FLAIR-3 volumes (Fig 2*A*). In the linear regression model, this resulted in a residual dispersion for large volumes (called "heteroscedasticity"), indicating that a linear relationship between FLAIR-2 and FLAIR-3 was not verified for the entire volume range (Fig 2*B*).

To take this dispersion into account, we applied a logarithmic transformation to volume measurements at TP2 and TP3. In line with this hypothesis, a strong linear relationship between log FLAIR-2 and log FLAIR-3 (Fig 2*C*) was observed. The residuals were distributed normally, with a high concen-

tration around zero and no major dispersion for large volumes (Fig 2*D*).

We next checked whether the association between FLAIR-2 (log of the volume) and FLAIR-3 (log of the volume) was still significant after the introduction of other variables (Table 1), which might modulate the apparent final infarct volume. After a backward stepwise selection of covariates, the final model amounted to a univariate relationship between log FLAIR-2 and log FLAIR-3 ($\beta = 0.91$, $P < .0001$, $n = 85$), with the R^2 coefficient equal to 0.81.

Relationship between Volumes and Clinical Outcomes

Figure 3 illustrates the associations between FLAIR volumes and mRS, which were rather similar by using either FLAIR-2 or FLAIR-3 volumes.

To evaluate the association between volumes and disabilities as well as the role of potential confounders more effectively, we performed a multivariable analysis predicting mRS dichotomized at a cutoff score of 1. Final models, including either FLAIR-2 or FLAIR-3 volumes, are shown in Table 4.
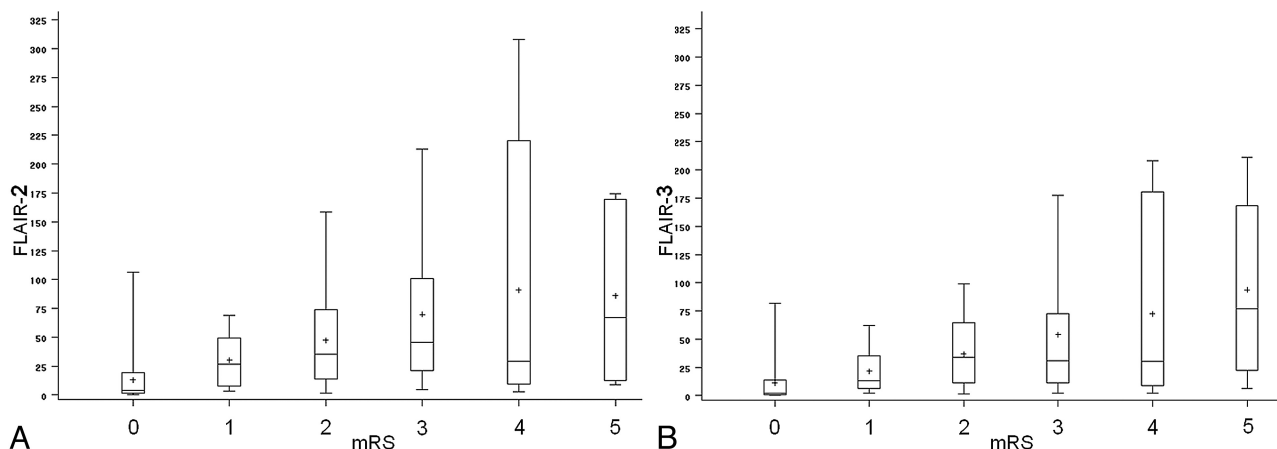
Fig 3. FLAIR volumes (either FLAIR-2 in part *A* or FLAIR-3 in part *B*) according to disability assessed by mRS. The cross inside the box represents the mean. The middle bar is the median. Upper and lower extremities of the box are Q1 and Q3, respectively. Maximum and minimum values are represented by extreme lines. Larger FLAIR volumes are associated with worse clinical prognosis with a similar tendency by using either FLAIR-2 or FLAIR-3.

Table 4: Final models of prediction of mRS ≤1 including either FLAIR-2 volume (model 1) or FLAIR-3 volume (model 2)[a]

Variables	OR	95% CI	P Value
Final model including FLAIR-2 volume (model 1)			
FLAIR-2 volume (per 10 cm^3 increase)	0.79	0.64–0.97	.026
Age			
Between 54 and 64 vs <54 years	0.05	0.01–0.32	.017
Between 64 and 74 vs <54 years	1.18	0.22–6.39	
Older than 74 vs <54 years	0.09	0.01–0.66	
Initial NIHSS score	0.82	0.72–0.93	.003
Reperfusion[b]			
Reperfusion ≥90% vs no abnormality	1.47	0.34–6.45	.005
Reperfusion <90% vs no abnormality	0.06	0.01–0.40	
Final model including FLAIR-3 volume (model 2)			
FLAIR-3 volume (per 10 cm^3 increase)	0.73	0.56–0.96	.023
Age			
Between 54 and 64 vs <54 years	0.06	0.01–0.47	.025
Between 64 and 74 vs <54 years	1.54	0.28–8.51	
Older than 74 vs <54 years	0.12	0.02–0.81	
Initial NIHSS score	0.82	0.71–0.93	.003
Reperfusion[b]			
Reperfusion ≥90% vs no abnormality	1.32	0.31–5.67	.008
Reperfusion <90% vs no abnormality	0.06	0.01–0.43	

[a] The Hosmer-Lemeshow test findings were nonsignificant (*P* = .213 and .969 for models 1 and 2, respectively), indicating a good model fit. The AUC was 0.897 for model 1 and 0.888 for model 2.
[b] If one changed the threshold from 1 to 2, the variable "persistent occlusion" appeared to be strongly related to the clinical outcome (OR, 0.08; 95% CI, 0.01–0.44; *P* = .004 for model 1; OR, 0.06; 95% CI, 0.01–0.36; *P* = .002 for model 2)

Age (introduced by class according to quartiles), initial NIHSS Score reperfusion, and FLAIR volumes (either FLAIR-2 or FLAIR-3) were independent predictors of mRS. However, the association of either FLAIR-2 or FLAIR-3 with disability was weak (OR, 0.79 and 0.73; 95% CI, 0.64–0.97 and 0.56–0.96; *P* = .026 and .023). The performances of the models were similar after introducing either FLAIR-2 (AUC = 0.897) or FLAIR-3 (AUC = 0.888) volumes.

When we changed the mRS threshold from 1 to 2, FLAIR volumes remained significantly associated with mRS (OR = 0.86 and 0.83; *P* = .002 for FLAIR-2 and FLAIR-3, respectively), in univariate analysis. When we used the mRS threshold at a cutoff of 2, multivariate models (FLAIR volume +

initial NIHSS Score + persistent occlusion) provided the same performance, introducing either FLAIR-2 (AUC = 0.870) or FLAIR-3 (AU = 0.887) volumes. With this cutoff at 2, the variable "persistent occlusion" was the stronger predictor, and FLAIR-2 and FLAIR-3 volumes did not reach significance, while FLAIR volumes were associated with the variable occlusion (median FLAIR-2 = 96.4 cm^3 versus 38.7 cm^3, *P* = .017; FLAIR-3 = 90.6 cm^3 versus 27.9 cm^3, *P* = .006 for patients with persistent occlusion versus patients without persistent occlusion).

Discussion

In this study, we found evidence for an early (<7 days) measurement of poststroke consequences because subacute volumes, even if they were overestimated by vasogenic edema, made it possible to accurately predict chronic (eg, fixed, after 30 days) volumes; subacute volumes provided the same estimation of clinical outcome as chronic volumes.

This finding could greatly improve the feasibility of phase II drug trials and proof-of-concept studies,[5,7] thereby facilitating the translation from preclinical studies in which infarct volume is traditionally used as the primary end point.[19] Nineteen percent of the variance of FLAIR-3 unexplained by FLAIR-2 (R^2 = 81%) could be considered as an acceptable loss of information when using subacute measurements. The advantages obtained would largely compensate for the loss of information: these are decreased study duration and costs, lower likelihood of confounding events unrelated to the intervention (such as another stroke), decreased loss to follow-up, and the lack of outcome measurements possibly leading to a biased and difficult interpretation of the study results. As an example, in our study, the outcome assessment at days 30–45 led to the loss of 5 patients, compared with only 1 patient at days 3–6; up to 30% of the patients were unable to undergo a 90-day MR imaging scan in the EPITHET study.[7] Earlier prognostic estimates can also assist with anticipating rehabilitation needs and can provide earlier information for patients and their relatives.

During the first week after insult, the apparent brain infarct, which is a dynamic process, has a tendency to expand, due to both vasogenic edema and increasing ischemic brain

tissue injury. The subsequent decrease may be explained by a combination of edema resolution, decrease of the inflammatory process, and atrophy.[9] Recently, Gaudinski et al[8] demonstrated that infarct volume did not vary significantly between days 30 and 90. This finding raised the question of whether it was possible to predict this fixed stage as early as days 3–6. Following this hypothesis, we found a good correlation between subacute and chronic volumes in line with previous work,[20,21] which nevertheless failed to take into account whether statistical conditions of a linear relationship were met for the full spectrum of volumes. Indeed, the estimation of FLAIR-3 was less efficient in the case of large volumes, resulting in a dispersion of the residuals. This could be accounted for by an evolution of the lesions, which was not uniform among patients, particularly in the case of large volumes in which early edema and later atrophy were predominant. We showed that a logarithmic transformation corrected this issue with a highly significant linear relation between log FLAIR-2 and log FLAIR-3, allowing us to conclude that the fixed infarct volume was predictable at days 3–6. Moreover, all covariates ceased to have an impact on volumes between TP2 and TP3, with most of the variability of FLAIR-3 explained by FLAIR-2 ($R^2 = 81\%$).

As a potential surrogate marker, infarct volumes must be correlates of the expected clinical outcomes.[4] Therefore, our study dealt with the relationship between volumes and clinical outcomes. We found that FLAIR volumes were significant and independent predictors of clinical outcomes. The correlations were not very strong, a point largely reported in the literature[22-24] and easy to understand, considering the numerous additional factors that can influence clinical outcomes at the disability/handicap level (On-line Fig 1).[25-27] Here, other clinical (initial stroke severity, age) or neuroimaging (recanalization, reperfusion) variables were significantly associated with the outcome, in line with previous studies.[25,28] The relatively weak association between infarct volumes and clinical outcomes does not necessarily discredit volume as a surrogate marker. In fact, treatment efficacy is certainly easier to demonstrate at this first hierarchical level (level of volumes in On-line Fig 1) in phase II studies, while it could be "diluted" at a higher levels (levels of disability or handicap in On-line Fig 1). An analogy could be made with multiple sclerosis. In this instance, the correlation between MR imaging–measured plaque volumes and clinical scales was even weaker, these volumes, nevertheless, being one of the most influential end points in neurologic pharmaceutical trials.[29]

More interesting than the radioclinical correlation per se, we evaluated the first direct comparison between volumes measured at different TPs and disability. The major result was that FLAIR-3 did not provide additional information because the strength of the association between FLAIR-3 and mRS was similar to that between FLAIR-2 and mRS. This was an additional argument in favor of not waiting longer than 1 week to assess final infarct volumes. This result was in line with a subanalysis from the EPITHET population, which found the same prognostic value for subacute or chronic volumes in predicting day 90 NIHSS scores.[21] Here, we extended these data by choosing mRS, a global measurement of disability and handicap widely used in phase III trials,[17,18] instead of NIHSS,

which is used to measure neurologic deficits and is not necessarily correlated with an individual's functional recovery.[16,30]

The dataset used in this study presents some limitations. First, reperfusion and recanalization were assessed at days 3–6, which is suboptimal for this end point because numerous patients exhibit spontaneous recanalization before this time span.[7] However, by choosing the "persistent occlusion" variable rather than recanalization in an effort to avoid merging early and late recanalization, we took into account this latter important factor. Second, disability scales were measured after 30–45 days, which may be too early because recovery only reaches a plateau 5–6 months after a stroke.[16] Functional and global scales are more valid in the long-term assessment of stroke consequences,[31] and this feature could have decreased the relationship between FLAIR volumes and mRS. Third, our patient sample was not representative of the full spectrum of patients with stroke, particularly because it included few severe strokes. This could limit the generalization of the obtained results to the entire stroke population. Nevertheless, similar characteristics, particularly initial NIHSS Score, can be found in most of the previous studies with multiple MR imaging[6] or in recent therapeutic trials.[18] Last, one must keep in mind that a correlation (between surrogate and clinical end points) is not sufficient for a surrogate to be valid.[4,32] Proper justification for such a replacement requires that the effect of the intervention on the surrogate end point predicts the effect on the clinical outcome.[4,32] It should be valid for recanalization therapy, such as fibrinolysis, which directly decreases the follow-up FLAIR lesion volume from the acute DWI-lesion volume.[33] More attention should be paid to other strategies like neuroprotectors or neurotrophic factors, which could influence disability or handicap levels, while the mean volume may remain unchanged.

Conclusions

Poststroke consequences could be assessed by FLAIR volume during the first week (days 3–6) rather than waiting for the chronic stage. The current study represents the first step in the complex process of the validation of MR imaging metrics for monitoring treatment effects, but it provides encouraging results and will have to be considered in future therapeutic trials, which could include a subacute infarct volume to assess other surrogacy criteria[4,32] before the definitive validation of this measure.

References

1. Phan TG, Donnan GA, Davis SM, et al. **Proof-of-principle phase II MRI studies in stroke: sample size estimates from dichotomous and continuous data.** *Stroke* 2006;37:2521–25

2. Warach S, Pettigrew LC, Dashe JF, et al. **Effect of citicoline on ischemic lesions as measured by diffusion-weighted magnetic resonance imaging: citicoline 010 investigators**. *Ann Neurol* 2000;48:713–22

3. Warach S, Kaufman D, Chiu D, et al. **Effect of the glycine antagonist gavestinel on cerebral infarcts in acute stroke patients, a randomized placebo-controlled trial: The Gain MRI Substudy.** *Cerebrovasc Dis* 2006;21:106–11. Epub 2005 Dec 9

4. Smith JJ, Sorensen AG, Thrall JH. **Biomarkers in imaging: realizing radiology's future.** *Radiology* 2003;227:633–38

5. Wintermark M, Albers GW, Alexandrov AV, et al. **Acute stroke imaging research roadmap.** *Stroke* 2008;39:1621–28

6. Albers GW, Thijs VN, Wechsler L, et al. **Magnetic resonance imaging profiles predict clinical response to early reperfusion: the diffusion and perfusion imaging evaluation for understanding stroke evolution (defuse) study.** *Ann Neurol* 2006;60:508–17

7. Davis SM, Donnan GA, Parsons MW, et al. **Effects of alteplase beyond 3 h after stroke in the echo-planar imaging thrombolytic evaluation trial (EPITHET): a placebo-controlled randomised trial.** *Lancet Neurol* 2008;7:299–309

8. Gaudinski MR, Henning EC, Miracle A, et al. **Establishing final infarct volume: stroke lesion evolution past 30 days is insignificant.** *Stroke* 2008;39:2765–68

9. Lansberg MG, O'Brien MW, Tong DC, et al. **Evolution of cerebral infarct volume assessed by diffusion-weighted magnetic resonance imaging.** *Arch Neurol* 2001;58:613–17

10. Tourdias T, Dousset V, Sibon I, et al. **Magnetization transfer imaging shows tissue abnormalities in the reversible penumbra.** *Stroke* 2007;38:3165–71

11. Renou P, Sibon I, Tourdias T, et al. **Reliability of the ECASS radiological classification of postthrombolysis brain haemorrhage: a comparison of CT and three MRI sequences.** *Cerebrovasc Dis* 2010;29:597–604. Epub 2010 Apr 23

12. Sibon I, Menegon P, Orgogozo JM, et al. **Inter- and intraobserver reliability of five MRI sequences in the evaluation of the final volume of cerebral infarct.** *J Magn Reson Imaging* 2009;29:1280–84

13. Sobesky J, Zaro Weber O, Lehnhardt FG, et al. **Which time-to-peak threshold best identifies penumbral flow? A comparison of perfusion-weighted magnetic resonance imaging and positron emission tomography in acute ischemic stroke.** *Stroke* 2004;35:2843–47

14. Neumann-Haefelin T, Wittsack HJ, Wenserski F, et al. **Diffusion- and perfusion-weighted MRI: the DWI/PWI mismatch region in acute stroke.** *Stroke* 1999;30:1591–97

15. Hosmer D, Lemeshow S. *Applied Logistic Regression.* 2nd ed. New York: John Wiley and Sons; 2000

16. Kasner SE. **Clinical interpretation and use of stroke scales.** *Lancet Neurol* 2006;5:603–12

17. **Tissue plasminogen activator for acute ischemic stroke: The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group.** *N Engl J Med* 1995;333:1581–87

18. Hacke W, Kaste M, Bluhmki E, et al. **Thrombolysis with alteplase 3 to 4.5 hours after acute ischemic stroke.** *N Engl J Med* 2008;359:1317–29

19. van der Worp HB, Kappelle LJ, Algra A, et al; the TESS Investigators; TESS II Investigators. **The effect of tirilazad mesylate on infarct volume of patients with acute ischemic stroke.** *Neurology* 2002;58:133–35

20. Beaulieu C, de Crespigny A, Tong DC, et al. **Longitudinal magnetic resonance imaging study of perfusion and diffusion in stroke: evolution of lesion volume and correlation with clinical outcome.** *Ann Neurol* 1999;46:568–78

21. Ebinger M, Christensen S, De Silva DA, et al. **Expediting MRI-based proof-of-concept stroke trials using an earlier imaging end point.** *Stroke* 2009;40:1353–58

22. Schiemanck SK, Post MW, Kwakkel G, et al. **Ischemic lesion volume correlates with long-term functional outcome and quality of life of middle cerebral artery stroke survivors.** *Restor Neurol Neurosci* 2005;23:257–63

23. Saver JL, Johnston KC, Homer D, et al. **Infarct volume as a surrogate or auxiliary outcome measure in ischemic stroke clinical trials: The RANTTAS Investigators.** *Stroke* 1999;30:293–98

24. Barrett KM, Ding YH, Wagner DP, et al; ASAP Investigators. **Change in diffusion-weighted imaging infarct volume predicts neurologic outcome at 90 days: results of the Acute Stroke Accurate Prediction (ASAP) trial serial imaging substudy.** *Stroke* 2009;40:2422–7. Epub 2009 May 14

25. Thijs VN, Lansberg MG, Beaulieu C, et al. **Is early ischemic lesion volume on diffusion-weighted imaging an independent predictor of stroke outcome? A multivariable analysis.** *Stroke* 2000;31:2597–602

26. Pillai JJ. **Insights into adult postlesional language cortical plasticity provided by cerebral blood oxygen level-dependent functional MR imaging.** *AJNR Am J Neuroradiol* 2010;31:990–96. Epub 2009 Dec 10

27. Cipriano LE, Steinberg ML, Gazelle GS, et al. **Comparing and predicting the costs and outcomes of patients with major and minor stroke using the Boston Acute Stroke Imaging Scale neuroimaging classification system.** *AJNR Am J Neuroradiol* 2009;30:703–09. Epub 2009 Jan 22

28. Rha JH, Saver JL. **The impact of recanalization on ischemic stroke outcome: a meta-analysis.** *Stroke* 2007;38:967–73

29. Sormani MP, Bruzzi P, Comi G, et al. **MRI metrics as surrogate markers for clinical relapse rate in relapsing-remitting MS patients.** *Neurology* 2002;58:417–21

30. Orgogozo JM. **The concepts of impairment, disability and handicap.** *Cerebrovasc Dis* 1994;4(suppl):2–6

31. Duncan PW, Jorgensen HS, Wade DT. **Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice.** *Stroke* 2000;31:1429–38

32. Prentice RL. **Surrogate endpoints in clinical trials: definition and operational criteria.** *Stat Med* 1989;8:431–40

33. Merino JG, Latour LL, Todd JW, et al. **Lesion volume change after treatment with tissue plasminogen activator can discriminate clinical responders from nonresponders.** *Stroke* 2007;38:2919–23

34. Rother J, Schellinger PD, Gass A, et al. **Effect of intravenous thrombolysis on MRI parameters and functional outcome in acute stroke <6 hours.** *Stroke* 2002;33:2438–45

35. Chalela JA, Kang DW, Luby M, et al. **Early magnetic resonance imaging findings in patients receiving tissue plasminogen activator predict outcome: insights into the pathophysiology of acute stroke in the thrombolysis era.** *Ann Neurol* 2004;55:105–12

36. Barber PA, Parsons MW, Desmond PM, et al. **The use of PWI and DWI measures in the design of "proof-of-concept" stroke trials.** *J Neuroimaging* 2004; 14:123–32