Visual and Region of Interest–Based Inter-Rater Agreement in the Assessment of the Diffusion-Weighted Imaging–Fluid-Attenuated Inversion Recovery Mismatch

Ivana Galinovic, MD, PhD; Josep Puig, MD; Lars Neeb, MD; Jorge Guibernau, MD; Andre Kemmling, MD; Susanne Siemonsen, MD; Salvador Pedraza, MD; Bastian Cheng, MD; Götz Thomalla, MD; Jens Fiehler, MD; Jochen B. Fiebach, MD

**Background and Purpose**—WAKE-UP is a randomized, placebo-controlled MRI-based trial of thrombolysis in wake-up stroke using the mismatch between a lesion’s visibility in diffusion-weighted imaging and fluid-attenuated inversion recovery (FLAIR) sequences as its main imaging inclusion criterion. Visual judgment of lesion conspicuity on FLAIR is however methodically limited by moderate inter-rater agreement. We therefore sought to improve rating homogeneity by incorporating quantitative signal intensity measurements.

**Methods**—One hundred forty-three data sets of patients with acute ischemic stroke were visually rated by 8 raters with respect to WAKE-UP study inclusion and exclusion criteria, and inter-rater agreement was calculated. A subanalysis was performed on 45 cases to determine a threshold value of relative signal intensity (rSI) between the ischemic lesion and contralateral healthy tissue which best corresponded to a visually established verdict of FLAIR positivity. The usefulness of this threshold in improving inter-rater agreement was evaluated in an additional sample of 50 patients.

**Results**—Inter-rater agreement for inclusion into the WAKE-UP trial was 73% with a free-marginal $\kappa$ of 0.46. A threshold of rSI which best correlated with the visual rating of lesions as FLAIR positive was 1.20. The addition of rSI measurements to visual evaluation did not change the inter-rater agreement.

**Conclusions**—Introducing a semiquantitative measure for FLAIR rSI did not improve the agreement between individual raters. However, enhancing visual assessment with rSI measurements can provide reassurance to local investigators in cases of uncertainty. (Stroke. 2014;45:1170-1172.)

**Key Words:** magnetic resonance imaging ● stroke

In patients with acute ischemic stroke, the combination of a lesion’s visibility in diffusion-weighted imaging (DWI) and in fluid-attenuated inversion recovery (FLAIR) imaging can offer information about time elapsed from stroke onset,1 with the so called DWI–FLAIR mismatch found to predict symptom onset <4.5 hours. However, visually judging lesion conspicuity on FLAIR is subjective and inter-rater agreements tend to be moderate;2–4 whereas an automated analysis is made complicated by low contrast and partial volume effects.3 The WAKE-UP trial4 relies on imaging criteria for the randomization of patients; in short, a DWI–FLAIR mismatch coupled with exclusion of hemorrhage and a DWI lesion involving less than one third of the middle cerebral artery territory. Because of the fact that this is a multicenter trial with images judged by local investigators, we wanted to inspect the reproducibility of our imaging criteria by assigning a large retrospective data set to 8 readers. This step, performed before the beginning of the trial, was meant to serve as refinement of both the imaging criteria and the methods for evaluating them.

**Materials and Methods**

This study was conducted using 143 patients’ images previously part of the PRE-FLAIR study1 sample. In the first step, these cases were visually rated by 6 raters (A.K., I.G., J.G., J.P., L.N., and S.S.) as well as by 2 senior neuroradiologists (J.B.F. and S.P.) according to the WAKE-UP study criteria.1 All raters received training in the form of official WAKE-UP material detailing the imaging criteria and showcasing a selection of typical problematic cases. Because a truly objective gold standard for the interpretation of lesion conspicuity on FLAIR does not exist, the consensus of the neuroradiologists was taken as the gold standard. In the second step, a subsample of 45 cases with a gold standard rating of either FLAIR negative or FLAIR positive was further processed by 5 raters (I.G., J.B.F., J.G., J.P., and L.N.) to determine signal intensity (SI) values of the acute ischemic lesion on FLAIR relative to the contralateral healthy tissue (rSI). This was done using the hot spot method (Figure 1), a method shown to give comparable results to the time-consuming method of obtaining...
Assessment of DWI–FLAIR Mismatch

Results

In the starting 143 cases, the percentage of agreement between junior raters for all imaging criteria was 57% with a $\kappa$ of 0.479. When only the final decision of inclusion or exclusion of patients (according to the WAKE-UP imaging criteria) was taken into account, the junior raters reached a percentage of agreement of 73% ($\kappa=0.455$) and the senior raters reached a percentage of agreement of 75% ($\kappa=0.50$). All cases of disagreement were subsequently decided on in a consensus read by the two neuroradiologists. A cutoff value of rSI which best correlated with this gold standard rating of lesions as FLAIR positive was 1.20, with true-positive rate values between 0.87 and 0.96 and false-positive rate values between 0.05 and 0.16 (Figure 2) This threshold was further used in the final sub-analysis; the results of which are presented in the Table. For 64% of cases, the decisions were identical across methods, for 22%, the rSI method showed improved inter-rater consensus, whereas in 14%, it showed reduced inter-rater consensus compared with visual assessment.

Discussion

Across several studies, interobserver agreement for judging lesion conspicuity on FLAIR has been shown as moderate.2–4 This is well explained by the gradual increase in SI on FLAIR after ischemic stroke. Exterior influences such as contrast and brightness, the personality and experience of the rater, as well as lesion size and image quality are also likely to play a role,1 as they probably did in our study as well. Agreement is higher for lesions which are clearly absent or present on the FLAIR image, but a significant fraction of patients present with a subtle FLAIR hyperintensity8 where individual rating becomes less predictable. In a study where judging FLAIR signal is the major criterion for randomization, assuring the same interpretation of FLAIR positivity across different centers is important. Of note, inter-rater agreement for this was, in our study, comparable with that reported for the identification of early ischemic signs on computed tomography.9 A subanalysis of the PRE-FLAIR

Table Interobserver Agreement and $\kappa$ for the Visibility of the DWI Lesion on FLAIR

<table>
<thead>
<tr>
<th>Agreement Rate %</th>
<th>$\kappa$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Multirater Analysis</td>
</tr>
<tr>
<td>Only visual assessment</td>
<td>72</td>
</tr>
<tr>
<td>Only rSI measurement</td>
<td>75</td>
</tr>
<tr>
<td>Combination*</td>
<td>73</td>
</tr>
</tbody>
</table>

*Here, the rater’s visual judgment was taken as the final verdict, unless he/she stated that an rSI measurement is needed in which case the rSI result was considered final.
study showed that a combination of visual and quantitative FLAIR analysis did not significantly improve the identification of patients within the 4.5-hour time window. In our study, semiquantitative measurements of rSI in manually placed hot spots achieved inter-rater agreements comparable but not superior to visual analysis. The combination of these 2 methods did not improve inter-rater agreement either. These findings are in line with previous results indicating that some aspects of DWI–FLAIR mismatch cannot be reliably quantified and remain subjective. In our study, the demand for a more objective measurement was signaled by the fact that in 18% to 38% of all cases even experienced raters reported a need for an rSI measurement to back up visual judgment. Assessing SI by placing hotspots is an intuitive and easy-to-apply method which is, in light of similar performances when compared with visual analysis, beneficial in reassuring investigators and whose advantages will be further assessed on prospectively collected data in the WAKE-UP study.

Sources of Funding
The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 278276 (WAKE-UP).

Disclosures
Dr Fiehler has received consultant or lecture fees from Bayer Healthcare, Boehringer Ingelheim, Boston Scientific, Bracco, Codman, Cordis, Covidien, ev3, Micrus, Microvention, Philips, Sanofi, Siemens, and Stryker. Dr Fiehler has received consultant or lecture fees from Boehringer Ingelheim, Lundbeck, Siemens, Syngis, and Synarc. Dr Pedraza has received consultant or lecture fees from Lundbeck and Synarc. The other authors report no conflicts.

References
Visual and Region of Interest–Based Inter-Rater Agreement in the Assessment of the Diffusion-Weighted Imaging– Fluid-Attenuated Inversion Recovery Mismatch
Ivana Galinovic, Josep Puig, Lars Neeb, Jorge Guibernau, Andre Kemmling, Susanne Siemonsen, Salvador Pedraza, Bastian Cheng, Götz Thomalla, Jens Fiehler and Jochen B. Fiebach

*Stroke*. 2014;45:1170-1172; originally published online February 20, 2014; doi: 10.1161/STROKEAHA.113.002661

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://stroke.ahajournals.org/content/45/4/1170

**Permissions:** Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

**Reprints:** Information about reprints can be found online at:
http://www.lww.com/reprints

**Subscriptions:** Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org/subscriptions/