CT and Diffusion-Weighted MR Imaging in Randomized Order

Diffusion-Weighted Imaging Results in Higher Accuracy and Lower Interrater Variability in the Diagnosis of Hyperacute Ischemic Stroke

J.B. Fiebach, MD; P.D. Schellinger, MD; O. Jansen, MD; M. Meyer; P. Wilde; J. Bender, MD; P. Schramm, MD; E. Jüttler, MD; J. Oehler, MD; M. Hartmann, MD; S. Hähnel, MD; M. Knauth, MD; W. Hacke, MD; K. Sartor, MD

Background and Purpose—Diffusion-weighted MRI (DWI) has become a commonly used imaging modality in stroke centers. The value of this method as a routine procedure is still being discussed. In previous studies, CT was always performed before DWI. Therefore, infarct progression could be a reason for the better result in DWI.

Methods—All hyperacute (<6 hours) stroke patients admitted to our emergency department with a National Institutes of Health Stroke Scale (NIHSS) score >3 were prospectively randomized for the order in which CT and MRI were performed. Five stroke experts and 4 residents blinded to clinical data judged stroke signs and lesion size on the images. To determine the interrater variability, we calculated κ values for both rating groups.

Results—A total of 50 patients with ischemic stroke and 4 patients with transient symptoms of acute stroke (median NIHSS score, 11; range, 3 to 27) were analyzed. Of the 50 patients, 55% were examined with DWI first. The mean delay from symptom onset until CT was 180 minutes; that from symptom onset until DWI was 189 minutes. The mean delay between DWI and CT was 30 minutes. The sensitivity of infarct detection by the experts was significantly better when based on DWI (CT/DWI, 61/91%). Accuracy was 91% when based on DWI (CT, 61%). Interrater variability of lesion detection was also significantly better for DWI (CT/DWI, κ=0.51/0.84). The assessment of lesion extent was less homogeneous on CT (CT/DWI, κ=0.38/0.62). The differences between the 2 modalities were stronger in the residents’ ratings (CT/DWI; sensitivity, 46/81%; κ=0.38/0.76).

Conclusions—CT and DWI performed with the same delay after onset of ischemic stroke resulted in significant differences in diagnostic accuracy. DWI gives good interrater homogeneity and has a substantially better sensitivity and accuracy than CT even if the raters have limited experience. (Stroke. 2002;33:2206-2210.)

Key Words: computed tomography ■ magnetic resonance imaging, diffusion-weighted ■ sensitivity and specificity ■ stroke, ischemic

Cerebral ischemia is the leading cause of the acute onset of severe neurological symptoms in stroke patients. Fibrinolysis has become an accepted treatment of cerebral ischemia during the first 3 hours after symptom onset and up to 6 hours in selected cases.1–4 The goal of fibrinolytic therapy is to recanalize an occluded basal brain artery and to prevent infarction in a hypoperfused tissue at risk.5 Before initiation of this treatment, however, other causes of acute symptoms such as seizure, encephalitis, and cerebral hemorrhage must be excluded. CT is the imaging modality most commonly used to exclude cerebral hemorrhage. It was also used in the European Cooperative Acute Stroke Study (ECASS) to exclude patients with extensive infarctions from recanalization therapy to reduce the incidence of intracranial hemorrhage (ICH).6 The main advantage of CT is its widespread availability in many hospitals 24 hours a day. The sensitivity of CT during the first 6 hours of cerebral ischemia was 64% in the ECASS reading panel with an accuracy of 67%. The local investigators of the ECASS trial reached only a 40% sensitivity (accuracy, 45%).7

According to the present literature, stroke MRI seems to be superior to CT as an imaging modality.8–11 All previous studies, however, lack 1 or more of the following: sufficient patient numbers, blinding of scan interpretation, relevant time window <6 hours, preselection of patients not based on the therapeutic target group, and most importantly a negative bias against CT, which always was performed before MRI with a mean time delay averaging 2 hours.12–14

In this study, we prospectively compared CT and diffusion-weighted MR imaging (DWI) in hyperacute stroke patients...
TABLE 1. Demographic Data and Delay Between Symptom Onset and Imaging

<table>
<thead>
<tr>
<th></th>
<th>CT First</th>
<th>MRI First</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, n (%)</td>
<td>24 (45)</td>
<td>30 (55)</td>
</tr>
<tr>
<td>Age, y</td>
<td>68.2</td>
<td>69.6</td>
</tr>
<tr>
<td>Median NIHSS score</td>
<td>11.5</td>
<td>10</td>
</tr>
<tr>
<td>Mean time from symptom onsets to CT, min</td>
<td>153.1</td>
<td>202.1</td>
</tr>
<tr>
<td>Mean time from symptom onsets to MRI, min</td>
<td>195.7</td>
<td>183.5</td>
</tr>
</tbody>
</table>

potentially eligible for thrombolytic therapy according to time window and baseline severity of stroke [National Institutes of Health Stroke Scale (NIHSS)]. To avoid any bias against CT resulting from delay after symptom onset, we prospectively randomized our patients to both imaging modalities (either stroke MRI or CT first followed by the other). The scans were read by neurologists and neuroradiologists with extensive experience in stroke imaging. We also investigated the accuracy of stroke imaging in the real-life clinical setting. Therefore, residents from the departments of neurology and neuroradiology read the images separately. Accuracy and sensitivity were calculated for each rater. Finally, we analyzed the interrater variability for both groups in both modalities.

Materials and Methods

Study Design

In a single-center, prospective trial conducted from February 2000 through March 2001, we screened 1320 patients with clinical signs of acute stroke, including transient ischemic attack (TIA). Inclusion criteria were anterior territory ischemic stroke within 6 hours of symptom onset with an NIHSS score of $>3$ and time interval between imaging modalities of $<90$ minutes. Baseline stroke severity was assessed with the NIHSS score. After clinical examination, patients were prospectively randomized for the sequence in which the imaging modalities were performed. Sealed randomization envelopes were kept in the neurological emergency ward. CT and DWI were performed by the junior radiology resident and/or the technician on call. Of the patients, 63% arrived outside the 6-hour time window (488 patients remaining), 48% had a TIA or NIHSS score of $<4$, and 18% had an ICH; some of these patients were randomized to imaging ($n=20$) but were not included into this study (166 patients remaining). Of these 166 patients, 29 had signs of posterior territory stroke; 13 did not give consent for randomization to imaging sequence; 10 had a time interval of $>90$ minutes between the 2 modalities; 27 were screened, imaged, and included in acute pharmacological trials; and 13 had MRI contraindications such as pacemakers, loss of vigilance, or recurrent vomiting. Twenty seemingly eligible patients were not included. For 11 of these patients, there was no adequate information available in our stroke data bank to explain why. In 9 theoretically eligible patients, MRI was not available because of maintenance procedures or nonavailability of the technician and/or physician (eg, because of a simultaneously performed neurointerventional procedure).

In total, images from 54 patients were presented to the readers: 50 with acute brain infarctions and 4 with symptoms of TIA. No images from studies before February 2000$^{11,14-16}$ were included to avoid any methodological pitfall according to changes in clinical setting (eg, dealing with stroke patients in the scanner increasingly becoming a routine procedure).

Results

In 54.5% of the study population, DWI was performed first. The mean delay between symptom onset and CT was 180...
minutes (SD, 85.3 minutes; range, 62 to 360 minutes). DWI was performed 189 minutes after symptom onset (SD, 83.0 minutes; range, 69 to 356 minutes). Mean delay between the 2 examinations was 29 minutes (SD, 17.8 minutes; range, 8 to 89 minutes). Demographic characteristics were similar in the 2 patient groups (Table 1). All images were of adequate diagnostic quality with typical variation according to patient compliance. All infarctions were in the MCA territory; 66% were left sided. The other 4 patients suffered from migraine, or postictal on early CT in 7 additional patients. The 4 CT scans of the early signs of infarction each resulted in an interrater variability of 61%. On the basis of early stroke signs, the interrater specificity of 65% (Table 2). The accuracy of lesion detection cases, the final decision (at least 4 readers) was negative. The patients with transient symptoms, migraine, or postictal gave

### Expert Ratings

All the experts reliably identified infarction on 20 CT scans. Early stroke signs were identified by at least 4 expert readers on early CT in 7 additional patients. The 4 CT scans of the patients with transient symptoms, migraine, or postictal paresis were judged normal by all experts. In 5 additional cases, the final decision (at least 4 readers) was negative. The mean CT sensitivity of the 5 experts was 61%, with a mean specificity of 65% (Table 2). The accuracy of lesion detection reached 61%. On the basis of early stroke signs, the interrater variability for identifying infarctions was \( \kappa = 0.51 \). Detecting the 4 early signs of infarction each resulted in an interrater variability of \( \kappa = 0.42 \) to 0.52; the estimation of lesion size, \( \kappa = 0.38 \). Differentiation between the various stroke types gave \( \kappa = 0.42 \) (Table 3).

On DWI, all of the experts identified infarcted sites in 40 cases. Infarction was identified by at least 4 observers on 7 additional DWI studies. One expert detected a hyperintense signal in a patient with normal findings on follow-up scan. The median DWI sensitivity of the experts was 91%, with a median specificity of 95%. An accuracy of 91% was found for lesion detection. The interrater variability was \( \kappa = 0.84 \) for detecting infarction in general and \( \kappa = 0.86 \) for identifying hyperintensities in the insular cortex. Rating of hyperintensities of the internal capsule or basal ganglia yielded \( \kappa = 0.62 \) and 0.60, respectively (Table 3). Agreement in stroke type identification gave \( \kappa = 0.52 \). The positive predictive value was excellent for both modalities (CT/DWI, 96%/100%). For CT, the negative predictive value was 12% (range, 9% to 17%); for DWI, 47% (range, 38% to 57%).

The agreement of stroke detection on DWI increased with the delay from symptom onset. For images made up to 120 minutes after ictus, \( \kappa = 0.76 \). It increased for the period between 120 and 240 minutes to \( \kappa = 0.79 \) and reached \( \kappa = 1 \) for the fifth and sixth hours after symptom onset. For CT, the values were \( \kappa = 0.52 \) (0 to 2 hours) and \( \kappa = 0.58 \) (3 to 4 hours) but only \( \kappa = 0.36 \) during the fifth and sixth hours after symptom onset. The major estimation of lesion size gave a clear result for 42 DWI and 26 CT. Findings were negative on 8 CT and 5 DWI studies. Lacunar lesions were not identified on any CT scans but were identified on 4 DWI images. Lesions smaller than one third of the MCA territory were detected on 15 CT scans and 25 DWI images.

Four infarcts were larger than one third and 1 was larger than two thirds of the MCA territory on CT. On DWI, 10 patients had large lesions. In 5 patients, the infarction was larger than one third, and in 3 additional patients, the lesion exceeded two thirds of the MCA territory, which were rated in agreement. In 2 additional cases, the raters judged discordantly whether the lesions were only larger than one third or larger than two thirds of the MCA territory.

### Novice Ratings

The novices reached a mean stroke sensitivity of 46% (range, 32% to 64%) on early CT, with a specificity of 56%. Accuracy ranged from 35% to 61% (mean, 46%). The negative predictive value differed by between 5% and 11% (mean, 7%). Mean positive predictive value was 93% (range, 90% to 96%). Detection of stroke signs showed a low agreement (\( \kappa = 0.38 \)). Estimation of lesion size was slightly more concordant (\( \kappa = 0.44 \)). In 12 patients, all novices detected signs of infarction in the same hemisphere. Eight CTs were concordantly judged as negative; 19 cases were rated the same by 3 of 4 raters.

Readings of the DWI yielded a sensitivity of 78% to 86% (mean, 81%). The agreement of lesion detection was high (\( \kappa = 0.76 \)). The novices reached a specificity of 100%. Mean positive predictive value was 100%. The estimation of lesion size was moderately concordant (\( \kappa = 0.55 \): 35 lesions were detected by all junior raters on DWI. They also judged 5 examinations in full agreement as negative.

### Table 2. Mean Statistical Results (Ranges) of the Ratings for Both Rating Groups

<table>
<thead>
<tr>
<th></th>
<th>Experts, %</th>
<th></th>
<th>Novices, %</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CT</td>
<td>DWI</td>
<td>CT</td>
<td>DWI</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>61 (52–70)</td>
<td>91 (88–94)</td>
<td>46 (32–64)</td>
<td>81 (78–86)</td>
</tr>
<tr>
<td>Specificity</td>
<td>65 (50–100)</td>
<td>95 (75–100)</td>
<td>56 (25–75)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Accuracy</td>
<td>61 (56–70)</td>
<td>91 (89–94)</td>
<td>46 (35–61)</td>
<td>82 (80–87)</td>
</tr>
<tr>
<td>Positive predictive</td>
<td>96 (94–100)</td>
<td>100 (98–100)</td>
<td>93 (90–96)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>Negative predictive</td>
<td>12 (9–17)</td>
<td>47 (38–57)</td>
<td>7 (5–11)</td>
<td>30 (27–36)</td>
</tr>
</tbody>
</table>

### Table 3. \( \kappa \) Values of Interrater Variability From Both Rating Groups

<table>
<thead>
<tr>
<th></th>
<th>( \kappa ), Experts</th>
<th>( \kappa ), Novices</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infarct detection</td>
<td>0.51</td>
<td>0.84</td>
</tr>
<tr>
<td>Infarct extent</td>
<td>0.38</td>
<td>0.62</td>
</tr>
<tr>
<td>Infarct type</td>
<td>0.42</td>
<td>0.52</td>
</tr>
<tr>
<td>Lesion of the insular cortex</td>
<td>0.46</td>
<td>0.86</td>
</tr>
<tr>
<td>Lesion of basal ganglia/internal capsule</td>
<td>0.52</td>
<td>0.78</td>
</tr>
</tbody>
</table>
Of the 5 lesions judged to be extensive infarctions on CT by the experts, the novices rated 25% as smaller. On DWI, 10 patients showed infarction larger than one third of the MCA territory. In 5%, the novices rated them falsely as smaller. The limited (less than one third of the MCA territory) infarctions identified by the experts on CT were judged as extended lesions in 13.3% by the novices. Based on DWI, the false-positive rate for extensive lesions was 3.4%.

Discussion

This study was designed to evaluate CT and DWI in hyperacute stroke under identical clinical conditions. Therefore, we randomized eligible patients suffering from hyperacute stroke with respect to the sequence of the imaging modalities. Powers and Zivin and Holloway have addressed the need for a routine means of comparing imaging modalities in ischemic stroke. The large number of patients screened and recruited in a 14-month period and the short delay between the 2 examinations indicate that we have established a routine procedure for dealing with those patients. However, some patients were not included in the study because of individual medical, infrastructural, or logistical reasons (eg, angiographic procedures with other patients during the hyperacute phase).

Lansberg and coworkers presented the results of a prospective trial of 19 cerebral infarctions with CT and DWI examination during the first 7 hours after stroke onset. They observed a higher accuracy of DWI for identifying acute infarction and a good sensitivity for detection of infarctions compared with CT. The median delay between CT and DWI was 2.5 hours, and CT was always done first. Therefore, the higher detection rate of DWI was based not only on a different imaging modality but also on the increase in edema over time. In a letter of correspondence, von Kummer and Gahn criticized the small number of patients in that study and the broad overlap of confidence intervals between the results of CT and DWI.

Sensitivity and accuracy of our CT rating (experts) were similar to the results of the experts’ rating according to von Kummer et al and as precise as the ECASS II CT reading panel. The assessment of the novices in our study showed a wide variety of sensitivity. The mean sensitivity was slightly higher than that of the ECASS local investigators. The experts’ ratings of early stroke signs were as moderately concordant as those observed by von Kummer and coworkers. The value of stroke diagnosis based on CT was also moderate. The judgment of lesion size in this study showed a fairly good agreement, beyond chance. The sensitivity of DWI was 91% by the expert raters. No rater reached a 100% detection rate as reported by Lansberg et al and González et al. Our recently published data from a former cohort of 31 patients with infarctions show values in lesion detection. The slightly lower value and sensitivity in the new cohort seem to reflect clinical reality in a routine setting. Even though DWI changes appear seconds after vessel occlusion in animal models and cortical hyperintensity was observed in patients 20 minutes after symptom onset, ischemia might not cause changes on DWI if the hypoperfusion allows structural metabolism. Furthermore, the lesions may also transitorily, partially, or permanently disappear on DWI.

The accuracy values in the 2 modalities prove the superiority of DWI. The interrater variability in infarct detection was excellent for DWI. In a larger cohort such as ours, the statistical result is slightly lower than reported in earlier studies. On the other hand, the statistical power with a total number of 54 patients seems to be sufficient. Furthermore, the interrater variability in detecting lesions is significantly better based on DWI.

Concerning the judgment of lesion size, DWI shows a better value than CT. In contrast to a previous study, we not only analyzed whether the lesion was larger than one third of the MCA territory but also distinguished 5 different categories. DWI gave a larger number of concordant ratings, and lacunar lesions were also detected more often. As observed in a previous study, the same lesions appear larger on DWI. However, before treatment of ischemic stroke can be considered, it is important to detect ischemic lesions as soon as possible and to estimate the potential infarct size because thrombolytic therapy in patients with large infarctions increases the rate of secondary, symptomatic ICH. Compared with the expert opinion as gold standard, the novices falsely judged large infarcts to be smaller than one third of the MCA territory in 25%, thus potentially giving recombinant tissue plasminogen activator (rtPA) to patients in whom it is contraindicated because of an excessive risk of ICH based on CT. This was substantially higher than if based on DWI (5%). Conversely, in 13.4% of the patients, the novices falsely judged a small infarction to be larger than one third of the MCA territory based on CT compared with only 3.4% if based on DWI. Therefore, novices are much more likely to either give rtPA to patients who should not be treated or withhold rtPA from patients who should be treated when their decision making is based on CT rather than MRI.

Based on DWI, the threshold of increased risk for symptomatic ICH may be larger than one third of the MCA territory. This hypothesis should be investigated in a multicenter study so that a sufficiently large number of patients is reached. In future studies, the Alberta Stroke Program Early CT Score (ASPECTS) would be a useful tool to describe the ischemic lesions in a more detailed way according to vascular anatomy. ASPECTS was established for CT within the first 3 hours of ischemic stroke and causes a higher agreement of different raters compared with the one third of the MCA rule. Before implemented into routine use, it should be validated for the 6-hour time window because efforts are being made to extend the rigid therapeutic time frame of 3 hours for thrombolytic therapy on the basis of improved imaging criteria.

Recently, several groups published results of reversible DWI and apparent diffusion coefficient changes that did not lead to infarction on follow-up images. Because these groups also investigated hyperacute stroke, this discrepancy between their data and ours cannot be explained by different clinical stages of ischemia. However, one difference between the results of those studies and ours with persistent lesions could be how the data were analyzed. We used a standardized presentation of the images with a low window level with a
narrow width. This means of presenting data could render the extent of a lesion smaller and show sharp margins between healthy tissue and the core of infarction.

One drawback of our study is that follow-up imaging was not available in 6 patients. All of these, however, had a persistent focal neurological deficit consistent with the initial diagnosis of ischemic stroke.

Comparison between our novices' results and the judgment of the local ECASS II investigators revealed a similar sensitivity. In daily routine, stroke patients should be treated as quickly as possible. In contrast to a study reading panel, no department is able to offer full expert assessment 24 hours a day, 7 days a week as part of clinical routine. Most patients are examined by interns or residents in emergency units nights and weekends. Under those circumstances the most reliable imaging modality should be used. When a daily stroke MRI routine is established, this modality leads to a short delay between the initial clinical examination and the initiation of therapy. Therefore, DWI should be used instead of CT as the modality of choice in stroke imaging. Despite the ongoing discussion about reversible changes, DWI shows the infarcted core in nearly every ischemic stroke patient and indicates the actual minimum extension of infarction at the time of imaging in nearly every patient.

References


CT and Diffusion-Weighted MR Imaging in Randomized Order: Diffusion-Weighted Imaging Results in Higher Accuracy and Lower Interrater Variability in the Diagnosis of Hyperacute Ischemic Stroke


Stroke. 2002;33:2206-2210
doi: 10.1161/01.STR.0000026864.20339.CB

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/33/9/2206