AJNR

Intraobserver and interobserver variability in schemes for estimating volume of brain lesions on MR images in multiple sclerosis.

M Filippi, M A Horsfield, M Rovaris, T Yousry, M A Rocca, C Baratti, S Bressi and G Comi

AJNR Am J Neuroradiol 1998, 19 (2) 239-244
http://www.ajnr.org/content/19/2/239

AJNR Am J Neuroradiol 19:239–244, February 1998

Intraobserver and Interobserver Variability in Schemes for Estimating Volume of Brain Lesions on MR Images in Multiple Sclerosis

Massimo Filippi, Mark A. Horsfield, Marco Rovaris, Tarek Yousry, Mara A. Rocca, Corrado Baratti, Sergio Bressi, and Giancarlo Comi

PURPOSE: **Our goal was to evaluate the intraobserver and interobserver reproducibility of measurements of brain lesion load in multiple sclerosis (MS) by using two proposed acquisition schemes.**

METHODS: **Three-millimeter-thick conventional spin-echo (CSE) and fast fluid-attenuated inversion-recovery (FLAIR) sequences were obtained and the lesions segmented using a semi-automated technique based on local thresholding to calculate intraobserver and interobserver reproducibility. These were compared with images obtained from two separate MR units in which 5-mm CSE sequences were obtained and segmented by using the local thresholding technique and also by manual outlining.**

RESULTS: **The intraobserver coefficient of variation was 4.0% (95% confidence interval [CI], 3.0% to 4.5%) for the 5-mm CSE sequence measured with manual outlining, 3.1% (95% CI, 2.5% to 3.2%) and 5.1% (95% CI, 4.1% to 5.6%) for the two sets of 5-mm CSE sequences measured using the local thresholding technique, 5.7% (95% CI, 3.9% to 6.6%) for the 3-mm CSE sequence, and 2.6% (95% CI, 2.1% to 2.7%) for the fast FLAIR sequence. The interobserver coefficient of variation was 7.1% (95% CI, 4.9% to 8.7%) and 8.3% (95% CI, 6.4% to 9.6%) for the two sets of 5-mm CSE sequences, 7.3% (95% CI, 4.7% to 9.1%) for the 3-mm CSE sequence, and 2.9% (95% CI, 2.3% to 3.3%) for the fast FLAIR sequence. The intraobserver and interobserver reproducibility of measurements obtained with the fast FLAIR technique was significantly better than those obtained with the other techniques.**

CONCLUSIONS: **Our data indicate that the intraobserver and interobserver variability in quantifying MS lesions can be reduced significantly with the use of fast FLAIR sequences, while no significant improvement is gained by reducing the section thickness from 5 mm to 3 mm.**

Change in T2 lesion load, as measured yearly using manual outlining on conventional spin-echo (CSE) magnetic resonance (MR) images, is used as a secondary end point in large-scale phase III clinical trials in multiple sclerosis (MS) (1, 2). Recently, however, image acquisition schemes that are more sensitive are starting to be used: CSE sequences with 3-mm-thick sections and fast fluid-attenuated inversion-recovery (FLAIR) sequences detected on average about 9% (3) and 18% (4) greater lesion volumes than did CSE sequences with 5-mm-thick sections. This was due to an increased detection rate both of small lesions and of lesions located in or near the cerebral cortex, as a consequence of better spatial resolution or image contrast. However, for clinical trials in MS, the intraobserver and interobserver reproducibility of lesion load measurements is more important than the sensitivity. Hence, several new approaches for segmenting MS lesions have been proposed, which, by reducing human interaction, increase the reproducibility of the results (1, 5–8). At present, natural history studies and clinical trials with these new MR imaging acquisition schemes are ongoing, but as yet there are no studies evaluating the reproducibility of lesion load measurements obtained with such

Received May 2, 1997; accepted after revision July 14.

Presented at the scientific meeting of the Society of Magnetic Resonance, New York, NY, April 1996.

Supported in part by a grant (ERBCHRXCT 940684) from the European Magnetic Resonance Network in Multiple Sclerosis (MAGNIMS) and by a grant (96/J/T49) of the Istituto Superiore di Sanità, National Ministry of Health, Rome, Italy.

From the Department of Neurology, Scientific Institute Ospedale San Raffaele, University of Milan (Italy) (M.F., M.R., M.A.R., C.B., S.B., G.C.); the Division of Medical Physics, University of Leicester (United Kingdom) (M.A.H.); and the Department of Diagnostic Radiology, Klinikum Grosshadern, University of Munich (Germany) (T.Y.).

Address reprint requests to Massimo Filippi, MD, Department of Neurology, Scientific Institute Ospedale San Raffaele, Via Olgettina 60, 20132 Milan, Italy.

schemes as compared with the present standard of reference (ie, lesion load measurements performed on CSE sequences with section thickness of 5 mm segmented by using manual outlining).

In this study, we evaluated the intraobserver and interobserver reproducibility of lesion load measurements in MS using 3-mm CSE sequences and 5-mm fast FLAIR sequences segmented using a semiautomated technique based on local thresholding. We then compared these results with those obtained using 5-mm CSE sequences, segmented with the same semiautomated technique and also by manual outlining.

Methods

Patients

Seventeen consecutive outpatients with clinically definite MS (9) were entered in the study. All had either relapsing-remitting or secondary-progressive MS. Patients with relapsing-remitting MS were defined as those having a clinical course characterized by acute exacerbations followed by complete or incomplete recovery (ie, leaving mild to moderate disabilities) and separated in time by periods of relative disease stability. Patients with secondary-progressive disease had an initial relapsing-remitting phase but became progressively disabled with or without superimposed relapses over at least the preceding 6 months. Written informed consent was obtained from all patients before they were admitted into the study. Ten patients were included in the first MR imaging protocol and seven in the second MR imaging protocol.

MR Imaging Protocols

First Protocol.—In center 1, MR imaging was performed with a 1.5-T unit. The pulse sequence used (spin-echo 2000/50 [repetition time/echo time]) provided a moderate T2 weighting and gave good definition of the MS lesions with some suppression of the cerebrospinal fluid signal. Twenty-four contiguous interleaved 5-mm-thick axial sections and 40 contiguous interleaved 3-mm-thick axial sections were obtained through the brain with a 22-cm field of view (FOV) and a 256 × 256 raw data and image matrix (acquisition time, 17 minutes 12 seconds for each section thickness). The images with 3-mm-thick sections had a poorer signal-to-noise ratio than did the 5-mm-thick images, but this was not to the detriment of lesion conspicuity.

Second Protocol.—In center 2, 18 contiguous interleaved 5-mm-thick axial CSE sequences (3000/20; acquisition time, 13 minutes 44 seconds) and fast FLAIR sequences (9000/150; inversion time, 2200; echo train length, 15; acquisition time, 5 minutes 6 seconds) were obtained on a 1.0-T unit; the FOV was 20 cm and the raw data and image matrix was 256 × 256 for both sequences.

The different acquisition parameters used for the 5-mm CSE sequences in the two centers reflected the usual practices at these institutions. Although the repetition times and echo times differed considerably, the resulting image contrasts were judged similar.

In both protocols, patients were positioned according to guidelines established by a European Community Committee for MS (10) and were not moved from the scanners for the duration of the imaging procedures. The patients were placed in a comfortable position at the center of the head coil using a standardized landmark and the indicator light. Then, planning scans (T1 weighted spin echo) were acquired as follows: 1) a single axial section was obtained; 2) from this, a coronal section was planned using an oblique projection if necessary to compensate for patient misalignment; 3) a sagittal section was prescribed from the coronal image, again compensating for any misalignment of patient position using the falx cerebri as a reference; 4) finally, the main series of sections was prescribed from the sagittal image. These sections ran parallel to a line that joins the most inferoanterior and inferoposterior parts of the corpus callosum.

Quantification of MR Imaging Abnormalities

Lesions were first identified and marked on the hard copies by a single observer. To evaluate the intraobserver agreements, another observer quantified the MR abnormalities on all images twice, using a quantitative semiautomated technique based on local thresholding (1, 5) on each of the two occasions. This observer also twice measured the lesion load present on the CSE images with section thickness of 5 mm obtained in center 1 using a manual tracing technique. The second set of evaluations was made 2 months after the first. The MR abnormalities were also quantified independently by two other observers using the same semiautomated technique in the same patient sample to calculate interobserver agreement. All the observers who quantified the lesion load present on the different images used the marked hard copies as a reference and were unaware of patient identity and clinical characteristics.

The software used for lesion volume measurements was the "/usr/image" library (University of North Carolina, Chapel Hill) and Dispim image display software (David Plummer, University College, London, United Kingdom) running on a Sun workstation (Sun Microsystems, Mountain View, Calif). The manual outlining measurements were performed by using a mouse-controlled cursor on the computer display. The cursor was moved to define the boundary of each lesion, and each outline was stored on a computer disk before automatic computation of the lesion volume. The measurements using the semiautomated local thresholding technique were performed using a mouse-controlled cursor on the computer display by clicking on the perimeter of the lesions. The computer program first examines the image in a region close to where the mouse was clicked to find the strongest local intensity gradient, which it considers to be the edge of the lesion. Then the lesion is outlined by following a contour of isointensity from this initial edge point, thus defining the lesion as a region of the image where the signal intensity is locally above the signal intensity at the initial edge position. This sometimes gave poor results because other structures (such as abutting gray matter) adjacent to the lesion were bright, leading to the contour moving away from the lesion outline. When this happened, the lesions were outlined manually. Both for manual outlining and for the local thresholding technique, the lesion volume was calculated simply as the lesion area multiplied by the section thickness.

Statistical Analysis

The intraobserver and interobserver agreements were calculated according to the statistical methods proposed by Bland and Altman (11).

Intraobserver Agreement.—All the lesion load assessments were transformed to an equivalent percentage scale of intraobserver agreement, according to the following formula:

$$\text{Intra observer agreement index} = 100 - \frac{|x_{1st} - x_{2nd}|}{(x_{1st} + x_{2nd})/2}$$

in which x_{1st} and x_{2nd} are measures obtained in twice-repeated evaluations using the same technique in the same patient.

Interobserver Agreement.—The lesion load assessments obtained by three observers using the three techniques were similarly transformed for each pair of observations to an equivalent percentage scale of interobserver agreement, according to the following formula:

$$\text{Inter observer agreement index} = 100 - \frac{|x_a - x_b|}{(x_a + x_b)/2}$$

in which x_a and x_b are the measures obtained by two different observers using the same technique in the same patient.

The measure of reproducibility used in this study was 2 standard deviations (SD) of the intraobserver and interobserver agreement indexes, according to the British Standards Institution (12). Therefore, the intraobserver coefficient of variation (COV) was equal to 2 SD of

$$\frac{|x_{1st} - x_{2nd}|}{(x_{1st} + x_{2nd})/2}$$

and the interobserver COV was equal to 2 SD of

$$\frac{|x_a - x_b|}{(x_a + x_b)/2},$$

where $|x_{1st} - x_{2nd}|$ is the absolute difference between two lesion volume measurements made by the same observer using the same technique, and $|x_a - x_b|$ is the absolute difference between the measures obtained by two different observers using the same technique. For the intra- and interobserver agreements, and also for the COV, a 95% confidence interval (CI) was calculated by using bootstrap analysis (1000 random samples were generated from the SAS package) with the algorithm described by Efron and Tibshirani (13).

To test differences between intraobserver and interobserver variabilities for the four techniques, Bartlett's test for homogeneity of variance was used. If the assumption of differences of variances between methods was significant, the appropriate univariate method was used to analyze mean differences (14). To evaluate the differences in variabilities and means in the comparison between intraobserver and interobserver agreement values, the statistical approaches described above were used.

Results

Intraobserver Agreement in Lesion Load Measurements

The median intraobserver agreement was 95.3% (CI, 94.6% to 95.6%) for the 5-mm CSE sequence measured with manual outlining. For the newer acquisition protocols, it ranged from 94.9% (CI, 94.1% to 96.5%) for the 5-mm CSE sequence obtained in center 1 to 98.5% (CI, 97.1% to 98.9%) for the fast FLAIR sequence (Table 1 and Fig 1). The intraobserver COV was 4.0% (CI, 3.0% to 4.5%) for the 5-mm CSE sequence measured with manual outlining. It ranged from 2.6% (CI, 2.1% to 2.7%) for the fast FLAIR sequence to 5.7% (CI, 3.9% to 6.6%) for the 3-mm CSE sequence (Table 1).

The intraobserver variabilities for the different techniques were not different. The comparisons between the means of intraobserver agreement corrected for differences of variance between techniques showed a significantly higher value for the fast FLAIR sequence as compared with the 5-mm CSE sequence measured with manual outlining (Welch analysis of variance [ANOVA]: F[1, 14.97] = 11.40;

TABLE 1: Median intraobserver agreements and coefficients of variation with 95% CI for MS lesion load measurements using different acquisition schemes and segmentation techniques

	CSE 5-mm, Manual Outlining	Local Thresholding Technique			
		CSE 5-mm (Center 1)	CSE 5-mm (Center 2)	CSE 3-mm	Fast FLAIR
Median agreement, % (CI)	95.3 (94.6–95.6)	94.9 (94.1–96.5)	98 (96.5–98.4)	96.8 (95.7–98.3)	98.5 (97.1–98.9)
COV, % (CI)	4.0 (3.0–4.5)	5.1 (4.1–5.6)	3.1 (2.5–3.2)	5.7 (3.9–6.6)	2.6 (2.1–2.7)

Note.—CSE indicates conventional spin-echo; FLAIR, fluid-attenuated inversion recovery; CI, 95% confidence interval calculated by using bootstrap analysis; and COV, coefficient of variation.

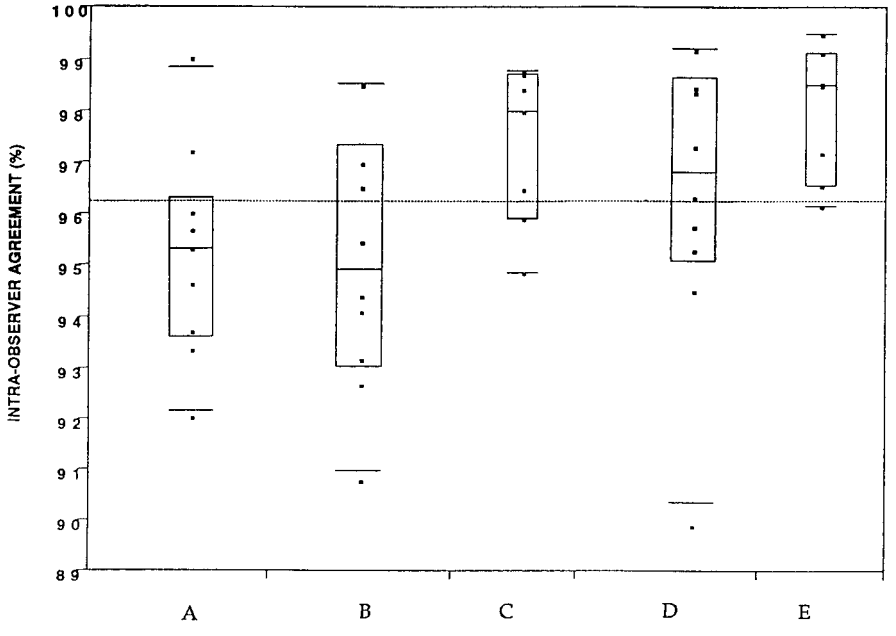

FIG 1. Median and quantile distributions of intraobserver agreements for the 5-mm CSE images segmented with manual outlining (*A*), the 5-mm CSE images obtained in center 1 (*B*) and in center 2 (*C*), and the 3-mm CSE images (*D*) and the fast FLAIR images (*E*) segmented using the local thresholding technique. The *rectangles* represent the 75%, 50%, and 25% quantiles, while the upper and lower *bars* represent the 90% and 10% quantiles, respectively. The *dotted line* represents the mean of the intraobserver agreement for the whole sample.

$P < .004$) and for the 5-mm CSE sequence obtained in center 2 and measured using the local thresholding technique and the 5-mm CSE sequence obtained in center 1 and measured with manual outlining (Welch ANOVA: F[1, 14.70] = 5.84; $P < .02$). No differences were found for any of the other comparisons.

Interobserver Agreement in Lesion Load Measurements

The median interobserver agreements ranged from 95.3% (CI, 94.8% to 96.4%) for the 5-mm CSE sequence obtained in center 2 to 97.6% (CI, 96.9% to 98.4%) for the fast-FLAIR sequence (Table 2 and Fig 2), and the interobserver COV ranged from 2.9% (CI, 2.3% to 3.3%) for the fast-FLAIR sequence to 8.3% (CI, 6.4% to 9.6%) for the 5-mm CSE sequence obtained in center 1 (Table 2).

The interobserver variability for the 5-mm CSE sequence was significantly higher than that obtained for the fast-FLAIR sequence (F[1, 40] = 13.98; $P < .0002$), but not different from that obtained for the 3-mm CSE sequence. The interobserver variability for the fast FLAIR sequence was also significantly lower than that for the 3-mm CSE sequence (F[1, 49] = 15.84; $P = .0001$).

The comparison between the means of interobserver agreement corrected for differences of variance between techniques showed a significantly higher value for the fast FLAIR sequence than for the 5-mm (Welch ANOVA: F[1, 26.40] = 8.11; $P = .008$) and 3-mm (Welch ANOVA: F[1, 40.62] = 7.10; $P = .01$) CSE sequences, whereas the difference between the 5-mm and 3-mm CSE sequences was not significant.

Comparisons between Interobserver Agreements Obtained with Different Techniques and Intraobserver Agreement Obtained by Measuring 5-mm CSE Sequences with Manual Outlining

Interobserver variability was significantly higher than intraobserver variability for measurements made with manual outlining on the 5-mm CSE images and for measurements obtained with the local thresholding technique on the 5-mm (F[1,38] = 5.5; $P < .01$) and 3-mm (F[1,38] = 3.86; $P < .05$) CSE images, but was not higher than that obtained for the fast FLAIR

TABLE 2: Median interobserver agreements and coefficients of variation with 95% CI for MS lesion load measurements using different acquisition schemes and segmentation techniques

	Local Thresholding Technique			
	CSE 5-mm (Center 1)	CSE 5-mm (Center 2)	CSE 3-mm	Fast FLAIR
Median agreement, % (CI)	96.1 (94.1–97.2)	95.3 (94.8–96.4)	96.7 (95.4–97.3)	97.6 (96.9–98.4)
COV, % (CI)	8.3 (6.4–9.6)	7.1 (4.9–8.7)	7.3 (4.7–9.1)	2.9 (2.3–3.3)

Note.—CSE indicates conventional spin-echo; FLAIR, fluid-attenuated inversion recovery; CI, 95% confidence interval calculated by using bootstrap analysis; and COV, coefficient of variation.

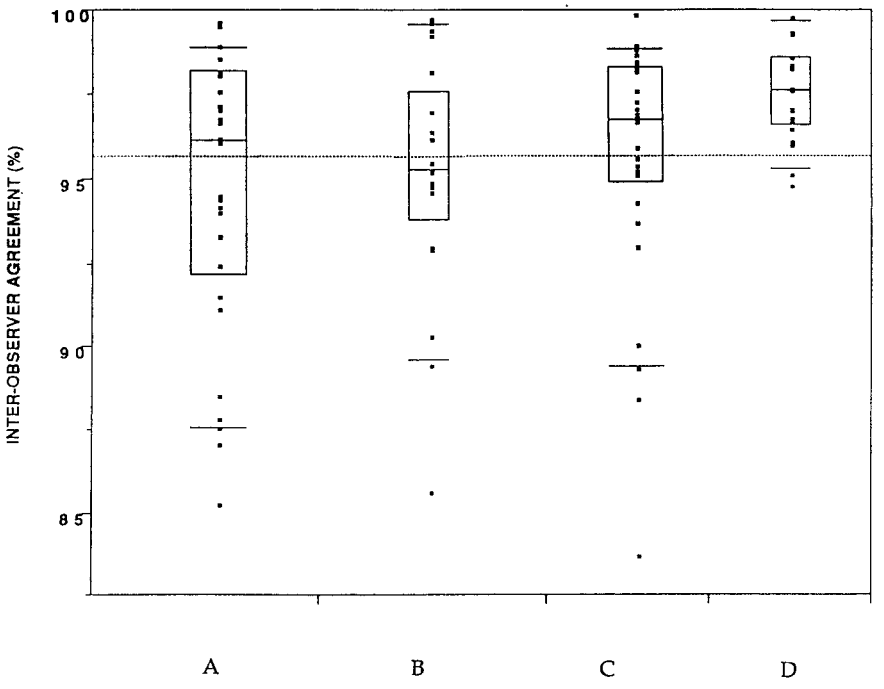

FIG 2. Median and quantile distributions of interobserver agreements for the 5-mm CSE images obtained in center 1 (*A*) and in center 2 (*B*), the 3-mm CSE images (*C*), and the fast FLAIR images (*D*) segmented by using the local thresholding technique. The *rectangles* represent the 75%, 50%, and 25% quantiles, while the upper and lower *bars* represent the 90% and 10% quantiles, respectively. The *dotted line* represents the mean of the interobserver agreement for the whole sample.

sequences. The mean of intraobserver agreement obtained by measuring the 5-mm CSE images with manual outlining differed significantly only from that obtained on the fast FLAIR sequences (Welch ANOVA: $F[1, 13.7] = 11.1$; $P = .005$). The mean of intraobserver agreement obtained for the 5-mm CSE images by using manual outlining did not differ from that of either the 3-mm or 5-mm CSE images obtained by using the local thresholding technique.

Discussion

Fast FLAIR and 3-mm CSE sequences depict a larger volume of lesions in patients with MS than do 5-mm CSE sequences (3, 4), which are currently used to monitor long-term disease evolution, both natural and modified by treatment (1, 2). Such sequences may be able to show smaller changes in lesion load over time, thus providing a more sensitive and complete assessment of MS lesion activity; however, the fundamental requirement of a technique used to assess lesion load in MS is that it should be reliable and reproducible over time, allowing clinical trials with adequate statistical power and relatively small patient populations. In recent years, several new partially or fully automated techniques for MS lesion segmentation have been proposed (1), which, by reducing human interaction, improve the reproducibility of the measurements. Different groups (5–8) have shown that by using intensity-based semiautomated techniques it is possible to improve significantly the reproducibility of lesion load measurements. With such segmentation techniques, any improvement in lesion contrast or resolution should result in further increases in the reproducibility of the measurements. Both fast FLAIR and 3-mm CSE sequences may be useful in this respect: fast FLAIR improves contrast by increasing the T2 weighting while suppressing the signal of cerebrospinal fluid; 3-mm CSE sequences give better definition of lesion edges by reducing partial-volume effects.

Our data indicate that the intraobserver and interobserver reproducibility in measuring MR lesion load in MS can be significantly improved by using fast FLAIR, while intraobserver and interobserver reproducibility is similar when the 3- and 5-mm CSE sequences are used. There are three possible explanations for this finding. First, lesion contrast on the fast FLAIR sequence we used is much higher than that on CSE sequences (4). It is therefore expected that a segmentation technique based on local intensity thresholding would work more reliably on such images. Second, to cover the entire brain with 3-mm-thick sections, more sections are needed. This leads to longer operator times for measurements (3), which may result in a reduction in reproducibility caused by operator fatigue. This is not the case for fast FLAIR, with its shorter analysis time (4). Third, it is conceivable that when more lesions and larger lesional volumes are present, the reproducibility of the results may be better, since errors with different directions (ie, overestimation and underestimation) may cancel each other out. Previous studies have shown that fast FLAIR images depict on average 18% more lesional volume than 5-mm CSE images (4), while the difference in lesional volume between 3- and 5-mm CSE images is only half that (ie, on average, about 9%) (3). It remains to be established whether FLAIR images with thinner sections, which have proved to be diagnostically useful in cases of suspected MS (15), will significantly affect the reproducibility of lesion load measurements.

In clinical practice, rapid acquisition relaxation enhancement (RARE) sequences (16), known also as fast spin-echo or turbo spin-echo are replacing CSE for the study of MS, since they enable dual-echo images to be obtained in a fraction of the time needed for CSE. Recent studies (17–19), however, found intraobserver and interobserver reproducibility of lesion load measurement to be similar, or slightly worse, for RARE sequences than for CSE. Two of these studies (18, 19) found the reproducibility of the measurements on fast FLAIR images to be better than that on RARE images. These results suggest that, when a reproducible measure of MR lesion load is needed in MS patients, fast FLAIR sequences give the best standard at present.

This study also demonstrates that the interobserver variability obtained by the combined use of fast FLAIR and the semiautomated technique based on local thresholding is lower than the intraobserver variability obtained by using 5-mm CSE and manual outlining, while this is not true for the semiautomated technique alone or when combined with 3-mm sections on CSE. This, coupled with the reduced acquisition and postprocessing times required by fast FLAIR (4), is clearly important for clinical trials in MS, in which it is common for hundreds of patients to be scanned several times and the use of a single observer for quantitative measurements might prove unrealistic. On the other hand, one can argue that standardization and optimization of fast FLAIR sequences in several centers involved in a clinical trial might be a challenging task as compared with that needed for 3-mm CSE sequences. It is indeed conceivable that the use of different MR units, with their possibly different magnetic field strengths, consequent changes in tissue relaxation times, and different implementations of imaging sequences, might all prove to be greater sources of inconsistency for fast FLAIR than those already demonstrated for CSE (20). In addition, the relationship between findings on fast FLAIR images and clinical aspects of the disease has not yet been established. Our results suggest that such clinical correlation is now needed.

Conclusions

Fast FLAIR, while a relatively new technique, gives a highly reproducible measure of the MS lesion burden. There are still questions about its sensitivity in different areas of the central nervous system and the clinical relevance of the extra disease burden seen on

FLAIR images, but we foresee its more widespread use in the monitoring of MS.

Acknowledgment

The Dispimage display program was written and provided by D. Plummer (Department of Medical Physics, University College, London, United Kingdom).

References

1. Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ, Miller DH. **Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis.** *Brain* 1995;118:1601–1612
2. Miller DH, Albert PS, Barkhof F, et al. **Guidelines for the use of magnetic resonance techniques in monitoring the treatment of multiple sclerosis.** *Ann Neurol* 1996;39:6–16
3. Filippi M, Horsfield MA, Campi A, Mammi S, Pereira C, Comi G. **Resolution-dependent estimates of lesion volumes in magnetic resonance imaging studies of the brain in multiple sclerosis.** *Ann Neurol* 1995;38:749–754
4. Filippi M, Yousry T, Baratti C, et al. **Quantitative assessment of MRI lesion load in multiple sclerosis: a comparison of conventional spin-echo with fast-fluid-attenuated inversion recovery.** *Brain* 1996;119:1349–1355
5. Filippi M, Horsfield MA, Bressi S, et al. **Intra- and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis: a comparison of techniques.** *Brain* 1995;118:1593–1600
6. Grimaud J, Lai M, Thorpe JW, et al. **Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques.** *Magn Reson Imaging* 1996;14:495–505
7. Paty DW, Li DBK, Oger JJF, et al. **Magnetic resonance imaging in the evaluation of clinical trials in multiple sclerosis.** *Ann Neurol* 1994;36:S95–S96
8. Van Walderveen MAA, Barkhof F, Hommes OR, et al. **Correlating MRI and clinical disease activity in multiple sclerosis: relevance of hypointense lesions on short TR/short TE (T1-weighted) spin-echo images.** *Neurology* 1995;45:1684–1690
9. Poser CM, Paty DW, Scheinberg L, et al. **New diagnostic criteria for multiple sclerosis: guidelines for research protocols.** *Ann Neurol* 1983;13:227–231
10. Miller DH, Barkhof F, Berry I, Kappos L, Scotti G, Thompson AJ. **Magnetic resonance imaging in monitoring the treatment of multiple sclerosis: concerted action guidelines.** *J Neurol Neurosurg Psychiatry* 1991;54:683–688
11. Bland JM, Altman DG. **Statistical methods of assessing agreement between methods of clinical measurements.** *Lancet* 1986;1:307–310
12. British Standards Institution. *Precision of Test Methods, I: Guide for the Determination and Reproducibility for a Standard Test Method. British Standard 5497*;1979
13. Efron B, Tibshirani R. **The bootstrap method for standard errors, confidence intervals and other measures of statistical accuracy.** *Stat Sci* 1986;1:1–35
14. Welch AB. **On the comparisons of several mean values: an alternative approach.** *Biometrika* 1951;38:330–336
15. Hashemi RH, Bradley WG, Chen DY, et al. **Suspected multiple sclerosis: MR imaging with a thin-section fast FLAIR pulse sequence.** *Radiology* 1995;196:505–510
16. Hennig J, Nauerth A, Friedburg H. **RARE imaging: a fast imaging method for clinical MR.** *Magn Reson Med* 1986;3:823–833
17. Rovaris M, Gawne-Cain ML, Wang L, Miller DH. **A comparison of conventional and fast spin-echo sequences for the measurement of lesion load in multiple sclerosis using a semiautomated contouring technique.** *Neuroradiology* 1997;39:161–165
18. Rovaris M, Yousry TA, Calori G, et al. **Sensitivity and reproducibility of fast-FLAIR, FSE, and TGSE sequences for the MRI assessment of brain lesion load in multiple sclerosis: a preliminary study.** *J Neuroimaging* 1997;7:98–102
19. Bastianello S, Bozzao A, Paolillo A, et al. **Fast spin-echo and fast fluid-attenuated inversion-recovery versus conventional spin-echo sequences for MR quantification of multiple sclerosis lesions.** *AJNR Am J Neuroradiol* 1997;18:699–704
20. Filippi M, van Waesberghe JH, Horsfield MA, et al. **Inter-scanner variation in brain MRI lesion load measurements in multiple sclerosis: implications for clinical trials.** *Neurology* 1997;49:371–377