

Intraobserver and Interobserver Variability in Measuring Changes in Lesion Volume on Serial Brain MR Images in Multiple Sclerosis

M. Filippi, M. Rovaris, M. P. Sormani, M. A. Horsfield, M. A. Rocca, R. Capra, F. Prandini, and G. Comi

PURPOSE: We evaluated the intraobserver and interobserver variability in measuring long-term changes in the volume of brain lesions on 5- and 3-mm-thick MR sections in patients with multiple sclerosis.

METHODS: Eighteen 18 patients were scanned on two separate occasions with a mean interval of 16.4 months between the two examinations. In each session, a scan with 24 contiguous 5-mm-thick axial sections and another with 40 contiguous 3-mm-thick axial sections was acquired consecutively without moving the patient. We assessed MR lesion load by using a semiautomated local thresholding technique.

RESULTS: Lesion volume was significantly higher on images with 3-mm-thick sections than on those with 5-mm-thick sections both at baseline and at follow up. Significant increases in total lesion volume were observed during the follow-up period on images obtained with both 5- and 3-mm-thick sections. The intra- and interobserver variability in measurements of changes in lesion volume was significantly higher on images with 5-mm-thick sections than on those with 3-mm-thick sections.

CONCLUSION: Our data indicate that the acquisition of thinner sections increases the reliability of the assessment of changes in brain lesion load on MR images in patients with multiple sclerosis.

Changes in lesion load on conventional T2-weighted MR images of the brain are used in virtually all long-term natural history studies and treatment trials as a measure of disease evolution in patients with multiple sclerosis (MS) (1, 2). Investigators have demonstrated that by decreasing the MR section thickness, the number and volume of brain abnormalities that can be detected increase significantly (3-5). Specifically, Filippi et al (3) showed in a cross-sectional study that the lesion volume detected on T2-

weighted images obtained with 3-mm-thick sections is on average about 9% greater than that on images obtained with 5-mm-thick sections, with a similar intraobserver reproducibility for repeated measurements (6). It has also been demonstrated that the use of 3-mm sections markedly reduces the variability of lesion load measurements as a result of small section repositioning errors (7) and the use of different scanners (8). For these reasons, there is a general consensus that brain MR images with 3-mm-thick sections should replace those with 5-mm-thick sections for monitoring patients with MS during clinical trials (2).

Traditionally, reproducibility of lesion load measurements has been evaluated by computing the abnormalities present on the same images two or more times (1). The variability in assessing the change in lesion load throughout a trial has not yet been investigated, although an understanding of this is important for power calculations when ascertaining the number of patients to enter into a trial (8). In our study, we evaluated and compared the intra- and interobserver variability in measuring changes in MS lesion load on serial brain MR images with two different section thicknesses (ie, 5 mm and 3 mm).

Received August 1, 1997; accepted after revision October 27.

Dr. Maria Pia Sormani is supported by a grant from TEVA Italy.

From the Department of Neurology, MS Biosignal Analysis Center, Scientific Institute Ospedale S. Raffaele, University of Milan, Milan, Italy (M.F., M.R., M.P.S., M.A.R., G.C.); the Unit of Epidemiology and Clinical Trials, National Institute for Cancer Research, Genoa, Italy (M.P.S.); the Division of Medical Physics, University of Leicester, Leicester, UK (M.A.H.); and the Departments of Neurology (R.C.) and Neuroradiology (M.P.), Spedali Civili, University of Brescia, Brescia, Italy.

Address reprint requests to Massimo Filippi, MD, Department of Neurology, MS Biosignal Analysis Center, Scientific Institute Ospedale San Raffaele, via Olgettina 60, 20132 Milan, Italy.

Methods

Patients

Eighteen consecutive patients (13 women, five men) with clinically definite MS (9) participated in the study. Thirteen had a relapsing-remitting course of the disease and five had a secondary-progressive course (10). The mean age was 33 years (SD, 7.9 years), the median duration of disease was 6 years (range, 2 to 15 years), and the median Expanded Disability Status Scale score (11) was 2.0 (range, 1.0 to 7.5). To be included, patients could not have had a clinical relapse (defined as the occurrence of a symptom or symptoms of neurologic dysfunction, with or without objective confirmation, lasting more than 24 hours) or a steroid treatment during the 3 months preceding the study. In addition, they could not have been treated with immunosuppressive or immunomodulating drugs for a period beginning 6 months prior to the study and lasting until the end of the study. For patients in whom a relapse occurred during the follow-up period, only steroid treatment was allowed (usually intravenous methylprednisolone 1 g/day for 5 days). The final MR studies were always obtained at least 3 months after the last relapse or steroid treatment. Written informed consent was obtained from all patients before inclusion in the study.

MR Imaging

Brain MR images were obtained on a 1.5-T unit. The pulse sequence used (spin-echo 2000/50/1 [TR/TE/excitation]) provided a moderate T2 weighting and gave good definition of the MS lesions, with some suppression of the CSF signal. During the same session, and without moving the patient from the scanner, two images were obtained: one with 24 contiguous interleaved 5-mm-thick axial sections and the other with 40 contiguous interleaved 3-mm-thick axial sections. For both sequences, the field of view was 22 cm, with a 256×256 matrix. For both 5- and 3-mm images, the blocks of sections covered the same brain region (ie, from the foramen magnum to the cerebral cortex). For all patients, MR examinations were performed both at entry into the study and at follow-up, with a mean interval of 16.4 months (range, 13 to 20 months) between the two studies. For both the entry and the follow-up studies, the MR sections were carefully positioned according to published guidelines (12). The patients were always placed in a comfortable position at the center of the head coil by using a standardized landmark and the indicator light. Planning images (T1-weighted spin-echo) were acquired in the following order: 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; and 4) 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.

MR Quantification

A single observer (T), unaware of when the MR study was performed, identified and marked the lesions on hard copies. Then, lesion volumes were measured on two different occasions (separated by 1 month), using the hard copies as a reference, by three independent technicians using a semi-automated local thresholding technique for lesion segmentation (1, 13).

Statistical Analysis

All the data were first logarithmically transformed in order to stabilize the variances (ie, to render the variance indepen-

Lesion volumes on entry and exit MR images (mean follow up period, 16.4 months) with 5- and 3-mm-thick sections in 18 patients with MS

Section Thickness	Mean \pm SE, mL	
	Lesion Volume on Entry Study	Lesion Volume on Exit Study
5mm	15.8 \pm 1.0	18.2 \pm 1.3
3mm	16.8 \pm 1.0	19.4 \pm 1.4

dent from the mean lesion volume). The changes in lesion volume were then assessed as ratios between follow-up lesion volume and baseline lesion volume. The variability was expressed as a proportional variation (ie, as a percentage of δV). Proportional variation is equal to $s - 1$, where s is the antilogarithm of the standard deviation of the transformed volumes. All the variances were evaluated by ANOVA testing. Differences between average volumes measured with the two different acquisitions (ie, 5- and 3-mm-thick sections) at the two time points (ie, baseline and follow up) were tested using a four-factors ANOVA model.

Results

Table reports the mean lesion volumes obtained from the two measurements performed by the three technicians for first and final images with 5- and 3-mm-thick sections. Lesion volumes were significantly higher on images with 3-mm-thick sections than on those with 5-mm-thick sections both at baseline (median difference, +5.6%; range, -6.5 to +46.5%; $P < .001$) and at follow up (median difference, +5.6%; range, -7.7 to +44.2%; $P < .001$). A median difference of +3.1% (range, -53.7% to +49.2%; $P .003$) in total lesion volume was observed over the follow-up period on images with 5-mm-thick sections, and a median difference of +4.5% (range, -46.2% to +41.9%; $P .003$) was found on those with 3-mm-thick sections.

The intraobserver proportional variation for measuring changes in lesion volume over time was 8.2% for images with 5-mm-thick sections and 3.9% for images with 3-mm-thick sections ($F_{57,57} = 4.16$; $P < .001$). The interobserver proportional variation for measuring changes in lesion volume over time was 8.4% for images with 5-mm-thick sections and 5.2% for images with 3-mm-thick sections ($F_{38,38} = 2.47$; $P < .05$).

Discussion

This study indicates that intra- and interobserver variability in measuring changes in lesion load on MR images of patients with MS is significantly reduced when 3-mm-thick sections are used rather than the conventional 5-mm thick sections. It also confirms previous studies in which it was found that a reduced section thickness led to detection of larger MS lesion volumes (3, 5). Because the percentage of increase in lesion volume was similar with both 3- and 5-mm section thicknesses, and because it is thought that images with 3-mm sections better depict changes in

small lesions, it follows that if new lesions developed during the follow-up period, they were not small, and that the rate of growth of existing lesions was independent of the size of the initial lesion(s).

A previous study (6) showed that the reproducibility of repeated measurements of the same images was similar for 3- and 5-mm-thick sections. In that study it was argued that the better lesion definition on 3-mm-thick sections was counterbalanced by increased operator fatigue when processing the larger lesion volumes and increased number of sections (3, 6). The present study, however, deals with the reproducibility of measuring MR lesion load changes over time, which is clearly more relevant in the context of MS clinical trials (2, 14). In this situation, the variability of the measures is not just a function of the observer but includes other factors, such as scanner differences (8) and scan repositioning accuracy (7).

The intra- and interobserver variability in the present study is similar and low, which might reflect the fact that the technicians who made the measurements were all trained at the same institution over a long period and had previously dealt with large numbers of images from clinical trials in MS. Nevertheless, the intra- and interobserver proportional variation for images with 3-mm-thick sections were about half those for images with 5-mm-thick sections. This is clearly important for clinical trials, in which measurement inaccuracies may mask treatment effects, since the median yearly increase of lesion load in MS, as suggested by the present study, is small. Since recent studies (6, 15–17) have suggested that measuring MS lesion load on rapid-acquisition relaxation-enhanced or fast fluid-attenuated inversion recovery images is reproducible, it is now timely to also evaluate the reproducibility of such sequences for measuring changes of lesion load over time. Definitive clinical trials are powered according to clinical end points and, for this reason, are usually overpowered for MR-derived measures (1, 18, 19). Thus, it is conceivable that in future clinical trials not all patients will undergo MR imaging, thereby reducing the cost of the trial and the inconvenience to the patients. In this case, although inpatient variability remains the largest source of variation for lesion load measurements in MS (20), any errors in the evaluation of lesion load must also be minimized.

Acknowledgments

We thank Manuela Beretta, Laura Legnani, and Paola Testa, who performed the MR measurements for this study.

References

- 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
- Evans AC, Frank JA, Antel J, Miller DH. **The role of MRI in clinical trials of multiple sclerosis: comparison of image processing techniques.** *Ann Neurol* 1997;41:125–132
- 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
- Filippi M, Yousry T, Horsfield MA, et al. **A high-resolution three-dimensional T1-weighted gradient echo sequence improves the detection of disease activity in multiple sclerosis.** *Ann Neurol* 1996;40:201–207
- Filippi M, Rocca MA, Horsfield MA, et al. **Increased spatial resolution using a 3-D T1-weighted gradient-echo sequence results in greater hypointense lesion volume in multiple sclerosis.** *AJNR Am J Neuroradiol* 1998;19:235–238
- Filippi M, Horsfield MA, Rovaris M, et al. **Intraobserver and interobserver variability of brain MR lesion volume estimation schemes in multiple sclerosis.** *AJNR Am J Neuroradiol* 1998;19:239–244
- Filippi M, Marciandò N, Capra R, et al. **The effect of imprecise repositioning on lesion volume measurements in patients with multiple sclerosis.** *Neurology* 1997;49:274–276
- Filippi M, JH van Waesberghe, MA Horsfield, et al. **Interscanner variation in brain MRI lesion load measurements in multiple sclerosis: implications for clinical trials.** *Neurology* 1997;49:371–377
- Poser CM, Paty DW, Scheinberg L, et al. **New diagnostic criteria for multiple sclerosis: guidelines for research protocols.** *Ann Neurol* 1983;13:227–231
- Lublin FD, Reingold SC, The National Multiple Sclerosis Society (USA) Advisory Board Committee on Clinical Trials of New Agents in Multiple Sclerosis. **Defining the clinical course of multiple sclerosis: results of an international survey.** *Neurology* 1996;46:907–911
- Kurtzke JF. **Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS).** *Neurology* 1983;33:1444–1452
- 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
- 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
- Barkhof F, Filippi M, Miller DH, Tofts PS, Kappos L, Thompson AJ. **Strategies for optimizing MRI techniques aimed at monitoring disease activity in multiple sclerosis treatment trials.** *J Neurol* 1997;244:76–84
- 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
- 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
- Rovaris M, Yousry T, Calori G, Fesl G, Voltz R, Filippi M. **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
- McFarland HF, Frank JA, Albert PS, et al. **Using gadolinium-enhanced magnetic resonance imaging lesions to monitor disease activity in multiple sclerosis.** *Ann Neurol* 1992;32:758–766
- Nauta JPP, Thompson AJ, Barkhof F, Miller DH. **Magnetic resonance imaging in monitoring the treatment of multiple sclerosis patients: statistical power of parallel-group and cross-over design.** *J Neurol Sci* 1994;122:6–14
- Stone LA, Albert PS, Smith ME, et al. **Changes in the amount of diseased white matter over time in patients with relapsing-remitting multiple sclerosis.** *Neurology* 1995;45:1808–1814