



Providing Choice & Value

Generic CT and MRI Contrast Agents



**FRESENIUS
KABI**

CONTACT REP

AJNR

Computerized method of lesion volume quantitation in multiple sclerosis: error of serial studies.

J H Simon, A Scherzinger, U Raff and X Li

AJNR Am J Neuroradiol 1997, 18 (3) 580-582

<http://www.ajnr.org/content/18/3/580>

This information is current as of July 29, 2025.

Computerized Method of Lesion Volume Quantitation in Multiple Sclerosis: Error of Serial Studies

Jack H. Simon, Ann Scherzinger, Ulrich Raff, and Xiaolin Li

Summary: This study was designed to evaluate a potentially important source of error in T2-hyperintense lesion measurement unique to longitudinal multiple sclerosis treatment trials that would not be detected by the standard intraobserver and interobserver error analyses. The effect of this "error of serial studies" was tested by using the standard-of-reference manual-outlining approach and a modified bi-feature space (statistical) approach applied to a database of five consecutive patients. To simulate the conditions of a longitudinal treatment trial, each patient had immediate repeat MR studies of the brain with imperfect head repositioning. The study hypothesis was confirmed that with an improved quantitative methodology, the "error of serial studies" (interseries error) would exceed the intraobserver error.

Index terms: Brain, magnetic resonance; Brain, measurements; Sclerosis, multiple

Change in the volume of the T2-hyperintense lesion load (T2-lesion) in multiple sclerosis (MS) has become an important secondary outcome measure in treatment trials and natural history studies (1–3). Previous evaluations of MS lesion quantitation methods have been tests of precision in which standard intraobserver and interobserver error analyses have been used. However, in longitudinal MS trials, study-to-study measurement error is compounded by several inconstant factors that affect the acquisition of magnetic resonance (MR) images, including change in the distribution of inhomogeneities of the magnetic and radio frequency fields and study-to-study variations in section position that cause partial volume-averaging effects. Apparent measurement changes occur that are unrelated to changes related to the disease and that are not detected with the usual evaluation criteria. The purpose of this study was to determine the effect of this special "error

of serial studies" by using a modified computerized segmentation procedure and the standard of reference method of manual outlining. We hypothesized that the interseries error would emerge as the dominant measurement error with an improved quantitative methodology.

Materials and Methods

Two consecutive MR imaging studies were acquired in five consecutive patients with MS. The data set included mean lesion volumes of 1, 11, 16, 18, and 68 cm³. The initial study (study A) was acquired with 3-mm interleaved sections using a spin-echo pulse sequence with parameters of 2133–2200/30,80/1 (repetition time/echo time/excitation) a 192 × 256 image matrix, 24 × 18–24-cm field of view, and flow compensation. Study B was identical, and was acquired immediately after study A after allowing a small change in head section position and scan angle. T2-lesion volume was determined by the standard of reference manual-outlining procedure (4) or by an in-house modification of a dual echo-based bi-feature space methodology (5).

First, raw imaging data were filtered by using an anisotropic diffusion filter (5). While viewing the filtered dual-echo images, the operator used a mouse to select a minimum of 40 training points (generally 60 to 100) within five tissue classes (gray matter, normal white matter, cerebrospinal fluid, background, and MS lesion) for every one to three adjacent sections. Each data set of one to three contiguous sections was partitioned into tissue classes by using the nonparametric Parzen window approach (6), creating "value-labeled" images. These were seeded for connectivity in areas determined prospectively as lesion by prior film review. Connectivity was manually broken with the use of a line tool when a predetermined lesion remained inseparable from nonlesional areas. One observer analyzed all studies, with duplicate measurements (measurements 1 and 2) on the same series repeated after a minimum interval of 1 day. Intraobserver (intraseries) error percentages were determined on the basis of the

Received January 31, 1996; accepted after revision May 3.

Presented in part at the annual meeting of the American Society of Neuroradiology, Chicago, Ill, April 1995.

Supported in part by a research grant from GE Medical Systems and DICYT, Universidad de Santiago de Chile (UR).

From the Department of Radiology, Box A-034, 4200 E Ninth Ave, University of Colorado Health Sciences Center, Denver, CO 80262. Address reprint requests to Jack H. Simon, MD.

AJNR 18:580–582, Mar 1997 0195-6108/97/1803-0580 © American Society of Neuroradiology

absolute value of the difference in measurements 1 and 2 divided by the mean. For interseries error, the first measurement of study A versus B and the second measurement of study A versus B were used for each subject. Analyses included comparison of means by paired *t* test and correlations using the Statistical Package for the Social Sciences (Chicago, Ill).

Results

For the manual-outlining approach, the intraobserver error ranged from 0.3% to 27.2% (mean, 11.0%; SD, 8.7), and the interseries error ranged from 2.0% to 34.6% (mean, 14.6%; SD, 9.0). Although larger, the increased mean interseries over intraobserver error was not significant ($P = .4$). For the modified bi-feature space approach, the intraobserver error ranged from 0% to 11.4% (mean, 4.4%; SD, 3.3), and the interseries error ranged from 1.7% to 27.4% (mean, 10.6%; SD, 8.7). This increase in interseries error as compared with intraobserver error was significant ($P = .04$). The intraobserver error was significantly less for the bi-feature space than for the manual-outlining approach ($P = .05$). Interseries differences in technique were not significant ($P = .4$). The correlation between absolute lesion volume based on the modified bi-feature method compared with the manual-outlining approach was high ($> .99$) and significant ($P < .001$). In most cases, the modified bi-feature space approach resulted in an apparently greater lesion volume than observed with the use of the manual-outlining approach, but there is no analytic accuracy standard with which to compare these methods.

Discussion

Quantitation errors in longitudinal MS treatment trials and natural history studies can be thought of as resulting from two major sources, including errors related to factors in the acquisition of MR studies and errors that are generated during the image analysis and quantitation procedures. The standard intraobserver and interobserver measures do not address these issues completely. In particular, intraobserver and interobserver analyses do not show the potential "error of serial studies" (interseries error) that we describe and observe in these simulations of a longitudinal MS trial, which can exceed in importance the intraobserver error.

Our results also suggest that improvements in computer-assisted analysis methods, such as the modified bi-feature space approach, can decrease technical measurement errors as compared with the standard or reference manual-outlining approach. The basis for this improvement was not explicitly tested in this study, but is believed to be primarily the result of a decrease in some of the arbitrary decisions about the borders of the MS lesions. Manual-outlining methods that use long repetition time, short echo time images rely on an observer's decision as to lesion boundaries, which are not always easily tracked, sharply defined, or of high-contrast. With the use of the modified bi-feature space approach, the mean intraobserver error for 3-mm-thick sections was about half that of the manual-outlining method. Nevertheless, the manual-outlining approach has been successfully used recently in the interferon beta-1b (IFN- β 1b) MS treatment trial (4) and will remain the standard of reference until improved methods are carefully validated in the setting of clinical trials or rigorous pretrial testing.

Despite the general agreement that the progression to multispectral (two or more input data sets) compared with simple visual or single echo-based "threshold" methods will improve the precision and accuracy in measuring MS lesions, a review of the literature suggests that the errors from computerized methods remain significant to date. Refinements in the more rapid computerized methods, however, are promising (5, 7, 8). The computerized approach to determining MS lesion volume described here comes at the expense of high total operator time, which is typically 1 to 4 hours per case, depending on the number of sections with pathologic areas per brain. We would argue, however, that the cost in human labor is small relative to the cost of a typical treatment trial, justifying for the near future the continued use of operator time-intensive procedures.

The minimal methodologic accuracy and precision required to detect treatment effects in MS clinical trials are not yet known. There is a general tendency toward increasing T2-lesion volume over a period of years in untreated patients with relapsing MS, estimated to be on the order of 5% to 7% per year, with a 30% 5-year increment observed in the IFN- β 1b trial (4). However, lesion volume may fluctuate considerably from month to month in active cases (9), far more than the methodologic errors we re-

port. As a result, a large sample size (several hundred patients) is required to detect treatment effects when T2-lesion volume is used as the outcome measure (4, 10). T2-lesion measurement error in longitudinal studies is a composite of the intraobserver error, the interobserver error (if more than one observer is used), and the "error of serial studies" (interseries error) that we describe. The effect of the interseries error is difficult to fully simulate and test, as changes in section position, radio frequency, and magnetic field, including magnetic susceptibility effects, all can in theory contribute to apparent lesion changes when pixel intensity is used as the basis for quantitation. Our analysis, with the partial simulation of change in section and angle, almost certainly will result in an underestimate of this interseries effect. Our preliminary data suggest that the error from repeat studies is likely, on average, to exceed the error measured by intraobserver criteria. It is, however, the interseries error that is most relevant to longitudinal treatment trials and natural history studies.

Conclusion

The "error of serial studies" may become a dominant source of error as computerized lesion measurement techniques improve. Although the magnitude of this effect in a large clinical trial is not determined by this pilot analysis, awareness of this phenomenon should improve understanding of its influence on outcome measures and guide development of future acquisition and analytic approaches.

Acknowledgments

We acknowledge the contribution of Charles Rush, Center for Human Simulation, University of Colorado Health Sciences Center, for modification of segmentation procedure software, and Robert Grossman, University of Pennsylvania, for helpful discussions concerning use of the bi-feature space methodology.

References

1. Miller DH, Albert PS, Barkhof F, et al. Guidelines for the use of magnetic resonance in monitoring the treatment of multiple sclerosis. *Ann Neurol* 1996;39:6-16
2. Simon JH. Magnetic resonance imaging of multiple sclerosis lesions: measuring outcome in treatment trials. *West J Med* 1996; 164:502-509
3. Whitaker JN, McFarland HF, Rudge P, Reingold SC. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Multiple Sclerosis* 1995;1:37-47
4. IFNB Multiple Sclerosis Study Group. Interferon beta-1b in the treatment of multiple sclerosis: final outcome of the randomized controlled trial. *Neurology* 1995;45:1277-1285
5. Kikinis R, Shenton ME, Gerig G, et al. Routine quantitative analysis of brain and cerebrospinal fluid spaces with MR imaging. *J Magn Reson Imaging* 1992;2:619-629
6. Duda RO, Hart PE. Nonparamagnetic techniques. In: *Pattern Classification and Scene Analysis*. New York, NY: Wiley; 1973:85-127
7. Jackson EF, Narayana PA, Falconer JC. Reproducibility of nonparamagnetic feature map segmentation for determination of normal human intracranial volumes with MR imaging data. *J Magn Reson Imaging* 1994;4:692-700
8. Mitchell JR, Karlik SJ, Lee DH, Fenster A. Computer-assisted identification and quantification of multiple sclerosis lesions in MR imaging volumes in the brain. *J Magn Reson Imaging* 1994;4: 197-208
9. 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
10. Jacobs LD, Cookfair DL, Rudick RA, et al. Intramuscular interferon beta 1-a for disease progression in relapsing multiple sclerosis. *Ann Neurol* 1996;39:285-294