AJNR

Whole-Brain Atrophy in Multiple Sclerosis Measured by Automated versus Semiautomated MR Imaging Segmentation

Jitendra Sharma, Michael P. Sanfilipo, Ralph H. B. Benedict, Bianca Weinstock-Guttman, Frederick E. Munschauer III and Rohit Bakshi

AJNR Am J Neuroradiol 2004, 25 (6) 985-996
http://www.ajnr.org/content/25/6/985

AJNR Am J Neuroradiol 25:985–996, June/July 2004

Whole-Brain Atrophy in Multiple Sclerosis Measured by Automated versus Semiautomated MR Imaging Segmentation

Jitendra Sharma, Michael P. Sanfilipo, Ralph H. B.Benedict, Bianca Weinstock-Guttman, Frederick E. Munschauer III, and Rohit Bakshi

BACKGROUND AND PURPOSE: **Semiautomated and automated methods are used to measure whole-brain atrophy in multiple sclerosis (MS), but their comparative reliability, sensitivity, and validity are unknown.**

METHODS: **Brain parenchymal fraction (BPF) was measured in patients with MS ($n = 52$) and healthy control subjects ($n = 17$) by four methods: semiautomated or automated segmentation and 2D or 3D pulse sequences. Linear measures of atrophy, whole-brain lesion volumes, and clinical data were used to explore validity.**

RESULTS: **The 2D automated method yielded unreliable segmentation and was discarded. The three other BPF methods produced data that were highly intercorrelated and indistinguishable by analysis of variance. In the MS group, semiautomated (2D: 0.84 ± 0.04, $P < .001$; 3D: 0.84 ± 0.05, $P = .04$) and automated 3D (0.83 ± 0.05, $P = .002$) BPFs were lower than controls (semiautomated 2D: 0.88 ± 0.02; 3D: 0.88 ± 0.03; automated 3D: 0.88 ± 0.03). In the MS group, the semiautomated ($r = -.79$ to $-.82$) and automated 3D ($r = -.81$) BPFs inversely correlated with third ventricular width and showed similarly robust correlations with the bicaudate ratio (all $r = -.74$). The semiautomated and automated BPFs showed similar, moderate correlations with T1 hypointense and FLAIR hyperintense lesion volume, physical disability (Expanded Disability Status Scale) score, and disease duration and similar differences between secondary progressive and relapsing-remitting patients. The intraobserver, interobserver, and test-retest reliability was somewhat higher for the automated than for the semiautomated methods.**

CONCLUSION: **These automated and semiautomated measures of whole-brain atrophy provided similar and nearly interchangeable data regarding MS. They discriminated MS from healthy individuals and showed similar relationships to established disease variables.**

Brain atrophy measurements are emerging as sensitive surrogate markers for therapeutic trials in multiple sclerosis (MS) (1). There is a growing body of evidence indicating that the MS disease process extends beyond overt multifocal white matter lesions shown by MR imaging to include disease in normal-appearing white (1, 2) and gray matter (3–5). The destructive nature has been revealed by the demonstration of axonal loss, Wallerian degeneration, and macroscopic brain and spinal cord atrophy occurring during early stages of MS (1, 6–20). Recent studies have shown that brain atrophy progresses significantly on an annual basis (1, 11–18) and bears a close relationship to clinical impairment (1, 11, 18–28). Thus, there has been increasing interest in developing methods for quantifying regional and whole-brain atrophy in MS (1, 28). Various measurement techniques have been used to measure whole-brain atrophy, including both automated and semiautomated algorithms (12, 13, 25, 26, 28–32), relying on either gross or normalized brain volume assessment. Automated methods have the potential advantage of efficiency and reproducibility (28), whereas semiautomated methods may lead to increased accuracy because of operator input (31).

To date, little work has been done to compare different data acquisition sequences and analysis pro-

Received June 20, 2003; accepted after revision December 18.

From the Buffalo Neuroimaging Analysis Center (J.S., M.P.S., R.H.B.B, R.B), the Jacobs Neurologic Institute (J.S., R.H.B.B., B.W.-G., F.E.M.III, R.B), Departments of Psychiatry and Psychology (R.H.B.B.) and Neurology (R.H.B.B., B.W.-G., F.E.M.III, R.B.), and Physicians Imaging Centers (F.E.M.III, R.B.), University at Buffalo, State University of New York, Buffalo, NY.

Address correspondence to Rohit Bakshi, MD, Center for Neurological Imaging, Brigham & Women's Hospital, Harvard Medical School, 221 Longwood Avenue, RF396, Boston, MA 02115.

cedures for measuring brain atrophy and to determine their relative precision, sensitivity, and reproducibility. Two published studies recently compared several automated and semiautomated methods and showed that cerebral atrophy measurements are affected by both segmentation algorithm and pulse sequence (26, 32). Another study comparing automated and semiautomated methods was reported in preliminary form, showing that semiautomated methods are more accurate than automated methods and are thus, perhaps, more valid because of the operator interaction assuring quality control throughout the algorithm (31). The goal of our study was to compare two software procedures used with two pulse sequences for their reproducibility and ability to detect whole-brain atrophy in patients with MS versus that in healthy controls. We also explored their validity as defined as the correlation between whole-brain atrophy and clinical measures of disease burden, linear measures of atrophy, and whole-brain MR imaging lesion assessments. The purpose of this study is to assess a few key methodologic issues underlying the assessment of whole-brain atrophy in MS to assist in the planning of clinical trials. This study should also be helpful in comparing studies in which different methodologies were used to measure whole-brain atrophy and in assessing the generalizability of any single study.

Methods

Subjects

We performed a cross-sectional study of 52 consecutively referred clinically confirmed patients with MS (38 women and 14 men) and 17 healthy controls (12 women and five men). The diagnosis of all patients was confirmed at a university-affiliated MS clinic. None of the patients with MS had other major clinical illnesses, were younger than 20 or older than 60 years, had used corticosteroids within 4 weeks preceding MR imaging, or had a history of substance abuse. Forty-three patients had relapsing-remitting MS, and nine had a secondary progressive clinical MS disease course. Physical disability was assessed by a single experienced neurologist (B.W.G.) blinded to the MR imaging findings by using the Expanded Disability Status Scale (EDSS; 33) within 1 week of MR imaging. EDSS scores in the MS group ranged from 1.0 to 8.0 (mean ± SD, 3.46 ± 1.90). The duration of MS ranged from 1 to 43 years (11.58 ± 8.7 years). Healthy volunteers recruited from the local community and hospital staff served as controls. Mean age of the MS group was 42.81 ± 8.5 (range, 23–61 years) and that of the control group was 35.94 ± 8.9 (range, 20–53 years). Analysis of variance (ANOVA) showed a significant difference between age of MS patients and that of healthy controls ($P = .006$). Therefore, all comparisons in this study were adjusted for age in the statistical analysis.

MR Imaging

MR Imaging Acquisition. Each subject underwent MR imaging performed by using the same protocol at a tertiary-care university facility (1.5-T Gyroscan ACS-NT; Phillips, Best, the Netherlands). Axial images of the brain were obtained, including 2D conventional spin-echo T1-weighted (TR/TE, 400/10), fast spin-echo T2-weighted (3000/120), and fast fluid-attenuated inversion recovery (FLAIR) images (TR/TE/TI, 8000/120/2200). The FLAIR imaging protocol has been presented in detail elsewhere (34). For all 2D axial sequences, matrix size was 192 × 256, number of signal averages was 2, number of sections was 24, field of view was 24 cm, and section thickness was 5 mm with no intersection gaps. All axial images were obtained in the canthomeatal plane by using internal landmarks. Coronal images of the whole brain were acquired with a high-spatial-resolution 3D gradient-echo technique (TR/TE, 24/7; flip angle, 30°; acquisition matrix, 256 × 256; sections, 70; section thickness, 2.5 mm; no section gap; field of view, 25 cm; and signals averaged, one), which resulted in a voxel size of 1.0 mm × 1.0 mm × 2.5 mm. Each component of MR imaging analysis was performed as detailed below by using a computer-assisted approach on a Sun Ultra 10 workstation (Sun Microsystems, Santa Clara, CA) by trained technicians (J.S., M.P.S.) who were unaware of clinical information.

MR Imaging Lesion Analysis. For analysis of hyperintense lesions on T2-weighted images, we used FLAIR findings, which have shown higher sensitivity and lower interobserver variability than do fast spin-echo T2-weighted images in the detection of areas of T2 prolongation in MS (35, 36). Our method of quantifying the whole-brain hyperintense lesion load has been recently detailed and validated (25). A masking and thresholding technique was used (Java Image software, version 1.0, Xinapse Systems, Northants, UK. All FLAIR axial sections from the midpoint of the cerebellum to the vertex were analyzed. Extracranial tissue was first removed by using a masking tool that involves an automated contour-tracing tool designed to isolate the brain contour. A second part of the masking procedure involves manually removing nonlesional extra-axial hyperintensities from within the brain surface contour, primarily the result of FLAIR artifacts in the CSF (34). A threshold was then applied to separate hyperintense lesions from those of nonlesional tissue. This constant threshold is reliable in separating the brain parenchyma and CSF for a given pulse sequence across subjects, despite the lack of performing signal intensity inhomogeneity correction; however, the threshold needs to be determined uniquely for different pulse sequences or different imaging platforms. The software then automatically calculates the whole-brain lesion load by multiplying lesion area by section thickness. Analysis of hypointense lesions on T1-weighted images, by using the methods described elsewhere (22, 25, 37), was performed by a single trained observer (J.S.) who was unaware of clinical details. In brief, this was performed by using a semiautomated edge finding and local thresholding technique (Java Image, version 1.0). Hypointensities on T1-weighted images were defined as lesions in the brain parenchyma that were of detectably lower signal intensity than those in white matter and were also at least partially hyperintense on FLAIR images. Whole-brain T1 hypointense lesion volume was then calculated automatically by the software program as the sum of the areas of all lesions multiplied by the section thickness.

MR Imaging Linear Atrophy Measures. To assess central atrophy, third ventricular width was measured from FLAIR images by using a method detailed elsewhere (14, 21, 37). A linear region of interest was drawn through the long axis of the third ventricle, parallel to the interhemispheric fissure in the section wherein the third ventricle was most visible. The width was measured by drawing a second linear region of interest perpendicular to the first at its midpoint, and recording its length. The bicaudate ratio, a marker of subcortical atrophy, was derived from FLAIR images, as detailed elsewhere (22). This is the ratio of the intercaudate distance to the brain width along the same line. The Java Image software was also used for these analyses.

Whole-Brain Atrophy MR Imaging Analysis. We used a normalized measure, the brain parenchymal fraction (BPF), defined as the ratio of brain parenchymal volume (tissue compartment) to the total brain volume within the surface contour (total intracranial volume). This was assessed by two segmen-

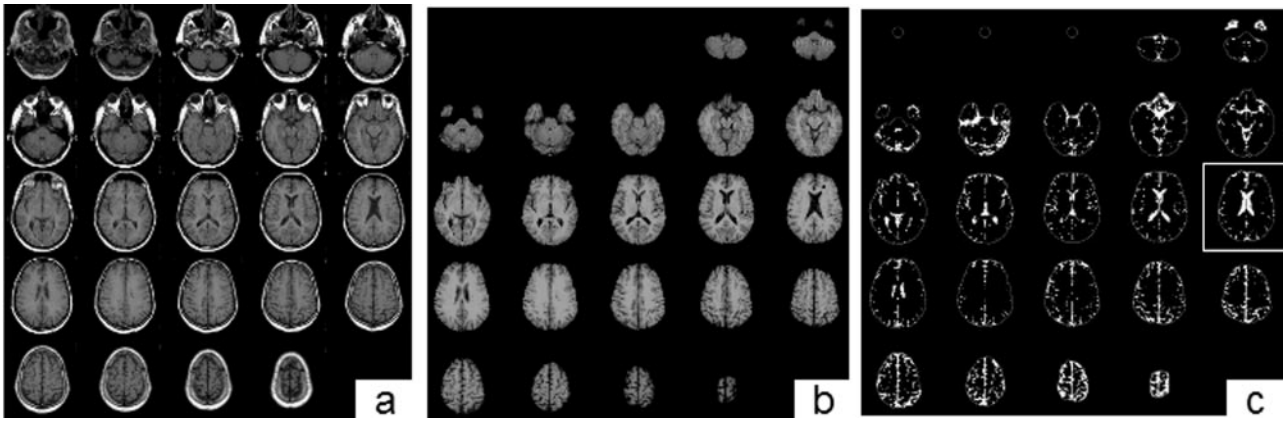

FIG 1. Semiautomated method of obtaining BPF. *A,* Raw T1-weighted 2D spin-echo noncontrast axial sequence. *B,* After masking (removal) of extracranial tissue. The segmented image (*C*) results from separation of the parenchyma (*black*) and CSF (*white*) into two compartments. The image surrounded by the white square (*C*) is used to identify normal-appearing white matter for the thresholding technique (see Methods). Adapted from Bermel et al (25).

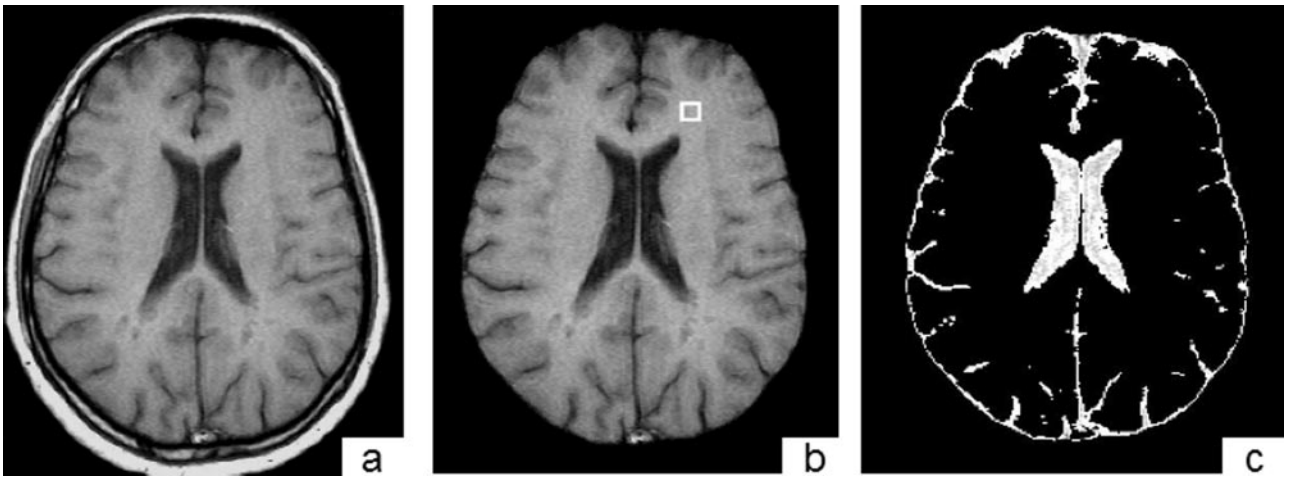

FIG 2. Semiautomated method of obtaining BPF. T1-weighted spin-echo 2D representative section from a patient with MS before masking (*A*), after masking the outer brain contour (*B*), and after thresholding (*C*). Panel *B* shows the region of interest in the normal-appearing white matter used for threshold determination. Adapted from Bermel et al (25).

tation techniques—semiautomated and automated—as described below.

Semiautomated Whole-Brain Atrophy Measurement. As shown in Figures 1 and 2, this BPF approach was based on a semiautomated computer-assisted technique by using masking and thresholding (Java Image, version 1.0), as detailed elsewhere (25). This was based on a previously described automated method (12) but was adapted to a semiautomated technique. We performed the semiautomated whole-brain atrophy measurement twice, on both the axial 2D T1-weighted and coronal 3D T1-weighted acquisitions. For the axial data set, all sections from the inferior cerebellum (where the vermis is first visible) to the top of the brain (up to the vertex) were analyzed. For the coronal data set, all the sections covering the brain were included; the inferior aspect of the brain was defined as the inferior extent of the cerebellar tonsils. These choices of sections and anatomic boundaries were designed to develop consistency in the brain volume sampled among subjects. The semiautomated method involved three major steps: 1) drawing the region of interest representing the external contour of the brain surface; 2) masking and removal of artifacts; and 3) segmentation of parenchyma from CSF. The extracranial tissue (eg, skull, orbits, skin, soft tissue) was removed from each section by using a semiautomated edge-finding tool included in the Java Image software package (Figs 1 and 2). The result was a single region of interest that included only the brain and CSF surrounded by a smoothly contoured brain surface. Manual correction was required when edge finding alone either included unnecessary tissue, excluded measurable parenchyma, or excluded subarachnoid space overlying the parenchyma. Artifacts present in the ventricles and the choroid plexus were manually added to the CSF compartment by setting the masking value to the CSF pixel value. CSF within the brain volume was then separated from the parenchyma by intensity thresholding (Figs 1 and 2). The thresholding step first required the placement of a 6×6-mm square region of interest in normal-appearing white matter on the axial 2D (Fig 2) or coronal 3D section wherein the full lateral ventricles were most visible. The mean pixel value was obtained for this region of interest and then was multiplied by 0.6 for the 2D and 0.5 for the 3D sections to define a constant threshold level. The determination of the optimal threshold thus depended on the pulse sequence and required development time through trial and error to separate the brain volume most effectively into data sets consisting of parenchyma and CSF pixels. The software program determined compartmental volumes by multiplying the total area of pixels by section thickness. Total operator-analysis time to determine the BPF by this semiautomated approach, including all masking and thresholding steps, was approximately 30 minutes per subject for the 2D sequences and 90 minutes for the 3D sequence because of the higher number of sections. With software improvements made after the com-

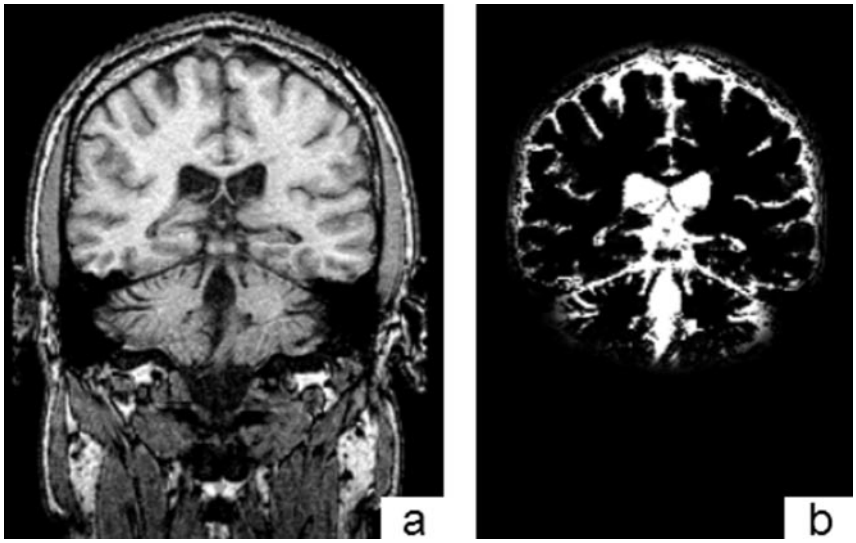

FIG 3. Automated method of obtaining BPF by using SPM-99 (see Methods). A T1-weighted 3D section showed the source image (*A*) and the resulting image after masking and segmentation into parenchyma or CSF (*B*).

pletion of the current study, however, the operator time has been reduced to 15 minutes (2D) and 45 minutes (3D). We did not perform signal intensity inhomogeneity correction in this method because, in a preliminary study, BPFs obtained from our semiautomated technique on raw (uncorrected) images were similar to those obtained from another semiautomated technique on inhomogeneity-corrected images (31). We specifically addressed the use of 24 sections with the 2D method instead of 28 sections that would have covered more of the brain. To evaluate this issue, we used the semiautomated method to measure BPF on 10 randomly chosen patients with MS by both the 24-section and 28-section data sets. The BPF measured by both techniques were similar, with a mean coefficient of variation (COV; SD/mean $\times 100\%$) of 0.4% (range, 0.2–0.5%).

Automated Whole-Brain Atrophy Measurement. BPF was determined by using SPM99 (Wellcome Department of Cognitive Neurology, Institute of Neurology, Queen Square, London) (38, 39) to align images in the same 3D orientation, correct for magnetic field inhomogeneity, and segment brain tissue into gray, white, and CSF compartments (Fig 3). This method of determining BPF was based on a previously published method (8). As for the semiautomated method, we performed the automated BPF measurement twice, on both the 2D T1-weighted axial and 3D T1-weighted coronal acquisitions. When aligning images, the origin was placed on the anterior commissure and the *x*-axis of the origin was centered to pass through the posterior commissure in the midsagittal plane. Brain images were aligned by matching the interhemispheric fissure to the origin's *x*- and *y*-axis in the anteroposterior and superior-inferior planes. The inferior and superior border of the brain was automatically selected by the brain mask. It included brain from the lower border of inferior cerebellum to the vertex. Realigned scans were masked to remove extracranial tissue (eg, skull, orbit, outer meningeal tissue) and were segmented into separate gray matter, white matter, and CSF images by using inhomogeneity correction (maximum level), which has been shown to improve brain image segmentation reproducibility (30). Final whole-brain volume measurements for each tissue compartment were based on the segmented, masked images from native brain images (Fig 3). The automated BPF performed on the 2D image resulted in poor segmentation, most likely because of the pulse-sequence limitations and large section thickness. Gross segmentation errors were common, such as the misclassification of CSF areas into the parenchymal compartment (data not shown). Thus, we dropped the automated 2D BPF from statistical analysis in this study.

Reliability and Variability of MR Imaging Assessments. Variability as an estimate of reliability and reproducibility was expressed as COVs. For the various BPF methods, intraobserver and interobserver reliability was determined from six subjects of the present study (three MS; three controls). To assess test-retest (scan-rescan) variability and stability of the BPF methods, two healthy volunteers (a 26-year-old man and a 36-year-old woman) underwent the MR imaging protocol twice (1 week apart). The intraobserver, interobserver, and test-retest COVs were 0.31%, 0.34%, and 0.41%, respectively, for the semiautomated 2D, 0.03%, 1.0%, 1.1%, respectively, for the semiautomated 3D, and 0.06%, 0.10%, and 0.10%, respectively, for the 3D automated BPF methods. The intraobserver and interobserver variability in the automated BPF technique was accounted for by differences in identification of the anterior commissure–posterior commissure line—the only step in the algorithm requiring operator input. As previously detailed, the intraobserver and interobserver COV were 1.7% and 4.5% for total T1 hypointense parenchymal lesion volume, 1.2% and 3.1% for total T2 hyperintense parenchymal lesion volume, 5.2% and 7% for third ventricular width, and 2.3% and 4.2% for bicaudate ratio (22, 25, 37).

Effect of Lesion Misclassification. Hypointense lesions due to MS can potentially contaminate whole-brain atrophy assessments on spin-echo or gradient-echo T1-weighted images because of misclassification of these lesions as CSF instead of as parenchyma (8, 38). We tested whether lesions affected BPF measurement in a subset of 10 patients from the MS group who had the highest T1-hypointense lesion burden. For the 2D semiautomated method, BPF was 0.803 ± 0.040 before and 0.804 ± 0.039 after correction of lesion misclassification (mean BPF change, 0.08%; COV, 0.06%). For the 3D automated method, BPF was 0.813 ± 0.041 before and 0.814 ± 0.041 after correction of lesion misclassification (mean BPF change, 0.19%; COV, 0.14%). Thus, the degree of contamination created by lesion misclassification was negligible for both methods, leading to a difference that was similar to or less than that seen because of scan-rescan effects.

Statistical Analysis

As noted above, there were MS versus normal differences in age, and BPFs were correlated with age ($P < .001$). Therefore, all analyses used to compare the relative relationships between independent and dependent variables controlled for the effects of age. Group comparisons were made by repeat-measures analysis of covariance controlling for age. Correlations were performed by using the Spearman rank correlation test for ordinal data and the Pearson test for continuous variables. Partial correlations were performed, where necessary, to adjust for age. Because of the number of statistical tests performed, a

TABLE 1: Comparison of methods of measurement of brain parenchymal fraction

		Semiautomated				Automated			
		2D		3D		2D		3D	
	No. of Cases	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Relapsing-remitting MS	43	0.85	0.04	0.85	0.04	NR	NR	0.84	0.05
Secondary progressive MS	9	0.81	0.04	0.80	0.03	NS	NR	0.80	0.04
All MS	52	0.84	0.04	0.84	0.05	NR	NR	0.83	0.05
Controls	17	0.88	0.02	0.88	0.03	NR	NR	0.88	0.03

Note.—2D and 3D sequences according to Methods section; MS, multiple sclerosis; NR, segmentation not reliable; SD, standard deviation; semiautomated and automated algorithms according to Methods section.

P value $< .01$ was considered to be statistically significant when interpreting statistical tests.

Results

The 2D automated method yielded unreliable segmentation and thus the data were discarded (see Methods). ANOVA comparing the three methods within the MS group and within controls did not reveal any significant difference. BPF values obtained by the three reliable methods in patients with MS and controls are shown in Table 1 and Figures 4 and 5. Two semiautomated BPFs (2D and 3D) were highly intercorrelated in the MS group ($r = 0.90$, $P < .001$), and the 2D semiautomated ($r = 0.87$, $P < .001$) and 3D semiautomated ($r = 0.91$, $P < .001$) BPFs were highly intercorrelated with the 3D automated BPF (Fig 4). Thus, for these three methods, the BPFs in the MS group were remarkably similar regardless of pulse sequence or segmentation algorithm. In controls, the range of values was restricted, and thus intercorrelations were highly positive but less robust; the semiautomated 2D BPF was intercorrelated with semiautomated 3D BPF ($r = 0.81$, $P < .001$), and the semiautomated 2D ($r = 0.83$, $P < .001$) and 3D ($r = 0.76$, $P = .001$) BPFs correlated with automated 3D BPF (Fig 4).

All three of the BPF methods were similarly sensitive in demonstrating whole-brain atrophy in the MS as compared with the control group. Analysis of covariance adjusted for age showed that the semiautomated (2D: 0.84 ± 0.04, range 0.71–0.91, $P < .001$; 3D: 0.84 ± 0.05, 0.72–0.92, $P = .04$) and automated 3D (0.83 ± 0.05, range 0.71–0.92, $P = .002$) BPFs were lower in the MS group than in healthy controls (semiautomated 2D: 0.88 ± 0.02, range 0.83–0.90; 3D: 0.88 ± 0.03, range 0.83–0.93; automated 3D: 0.88 ± 0.03, range 0.80–0.91) (Table 1, Fig 5), consistent with whole-brain atrophy in the MS group. The three BPFs showed a similar and higher degree of atrophy in secondary progressive (semiautomated BPFs: 2D: 0.81 ± 0.04, $P = .04$; 3D: 0.80 ± 0.03, $P = .03$; automated 3D: 0.80 ± 0.04, $P = .09$) versus relapsing-remitting patients with MS (semiautomated BPF: 2D: 0.85 ± 0.04; 3D: 0.85 ± 0.04; automated 3D: 0.84 ± 0.05) (Table 1, Fig 5), with the differences approaching significance.

In the MS group, the semiautomated (2D: $r = -82$, $P < .001$; 3D: $r = -.79$, $P < .001$) and automated 3D ($r = -.81$, $P < .001$) BPFs showed strikingly similar robust inverse correlations with third ventricular width (Table 2, Fig 6), which indicates a relationship between whole-brain atrophy and central brain atrophy. Similarly, semiautomated (2D: $r = -.74$, $P < .001$; 3D: $r = -.74$, $P < .001$) and automated (3D: $r = -.74$, $P < .001$) BPFs showed identical robust inverse correlations with bicaudate ratio (Table 2, Fig 7), which indicates an association between whole-brain atrophy and subcortical brain atrophy. The semiautomated (2D: $r = -.38$, $P = .006$; 3D: $r = -.44$, $P = .001$) and automated (3D: $r = .48$, $P < .001$) BPFs showed similar moderate inverse correlations with whole-brain T1 hypointense lesion volume (Table 2; Fig 8). The automated 3D BPF ($r = -.44$, $P = .001$) showed a somewhat higher correlation with total brain FLAIR hyperintense lesion volume than did the semiautomated BPFs (2D: $r = -.24$, $P = .09$; 3D: $r = -.35$, $P = .01$) (Table 2; Fig 9). The semiautomated (2D: $r = -.44$, $P <. 001$; 3D: $r = -.47$, $P < .001$) and automated 3D ($r = -.33$, $P = .008$) BPFs showed similarly moderate inverse correlations with physical disability (EDSS score; Table 2). The semiautomated (2D: $r = -.50$, $P < .001$; 3D: $r = -.39$, $P = .002$) and automated 3D ($r = -.53$, $P < .001$) BPFs showed similar moderate inverse correlations with disease duration (Table 2).

Discussion

In the present study, we compared two segmentation algorithms (semiautomated versus automated) and two pulse sequences (T1-weighted 5-mm conventional spin-echo 2D and T1-weighted 2.5-mm gradient echo 3D) for the measurement of normalized whole-brain atrophy in MS. Both approaches were performed by using commercially available software packages. To use whole-brain atrophy as a surrogate marker for disease progression and for assessing effects of therapies, it is desirable that segmentation techniques have validity, accuracy, reproducibility, and sensitivity. It has been presumed that automated measurement of whole-brain atrophy is preferred because of inherent efficiency and reproducibility (28). Semiautomated measurements, however, have the advantage of operator interaction to ensure accuracy of the masking and segmentation steps (25, 31). Our study indicates that there are many similarities among the methods. Both semiautomated and automated

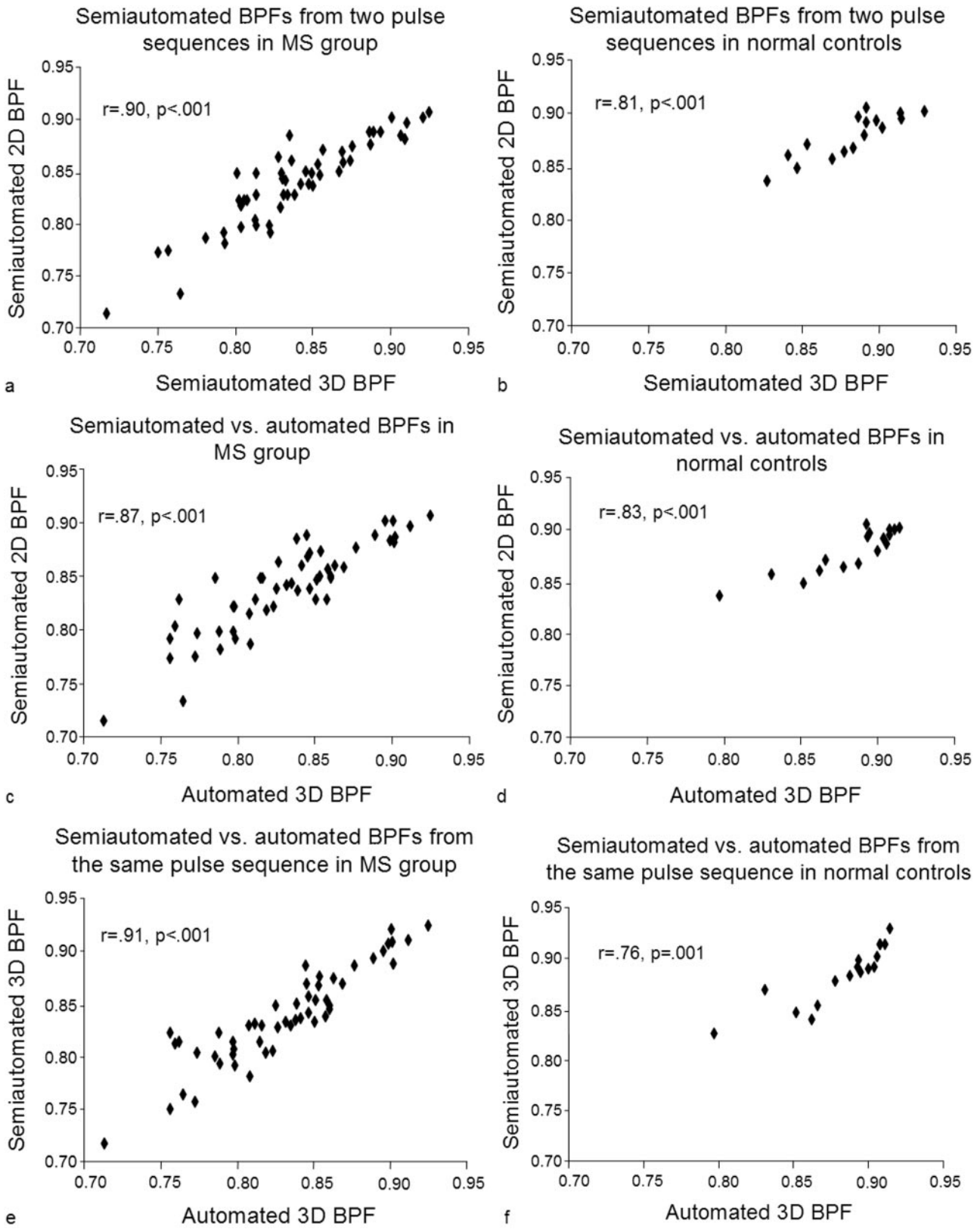

FIG 4. Scatterplots of semiautomated-2D versus semiautomated-3D BPF in the MS (*A*) and control (*B*) groups. Semiautomated 2D versus automated 3D BPF in the MS (*C*) and control (*D*) groups. Semiautomated 3D versus automated 3D BPF in the MS (*E*) and control (*F*) groups. BPFs derived by the two methods were highly correlated within the MS group and control group. The lower intercorrelation in the control versus MS group in all three methods is most likely related to restricted range.

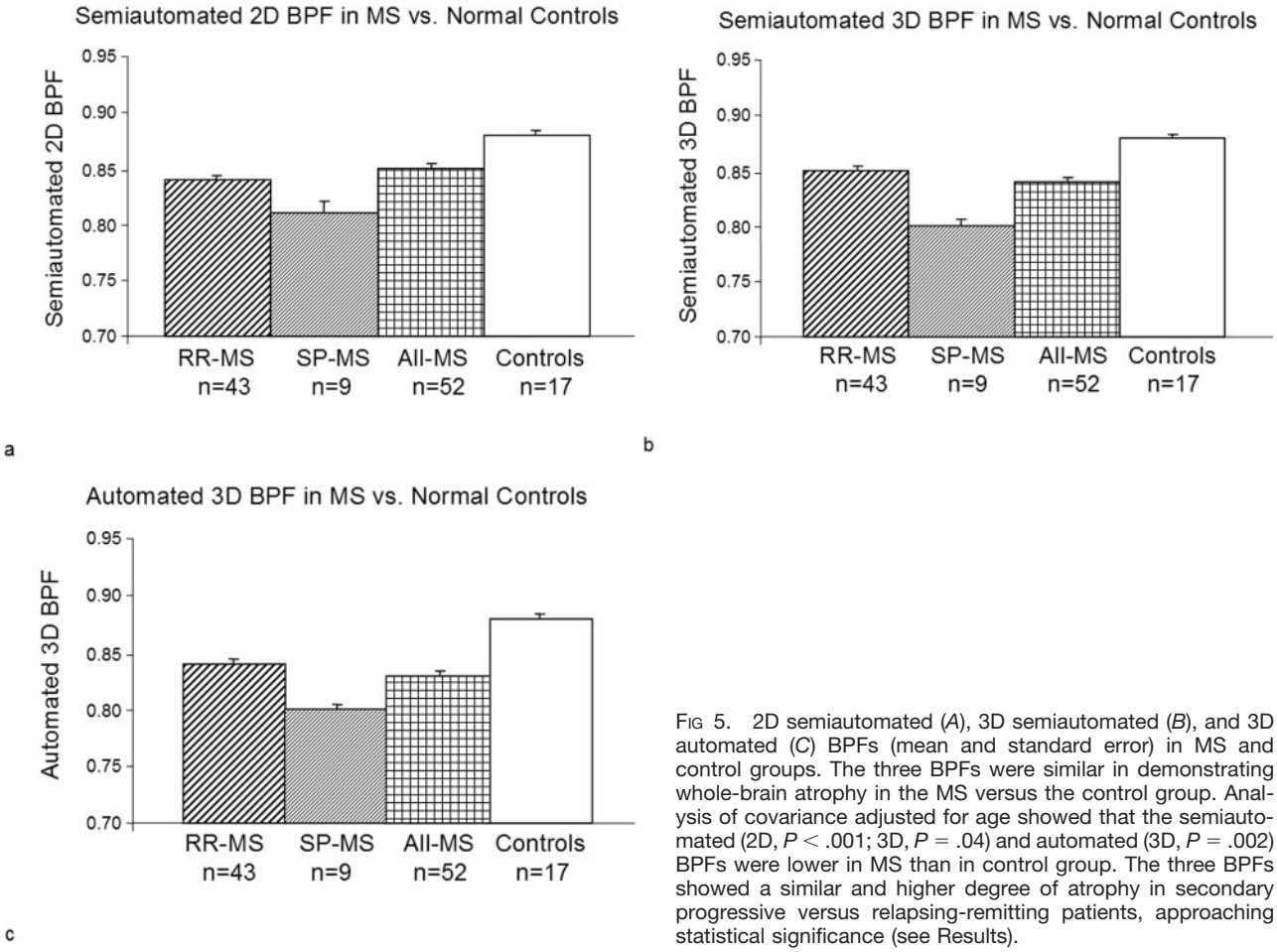

FIG 5. 2D semiautomated (*A*), 3D semiautomated (*B*), and 3D automated (*C*) BPFs (mean and standard error) in MS and control groups. The three BPFs were similar in demonstrating whole-brain atrophy in the MS versus the control group. Analysis of covariance adjusted for age showed that the semiautomated (2D, $P < .001$; 3D, $P = .04$) and automated (3D, $P = .002$) BPFs were lower in MS than in control group. The three BPFs showed a similar and higher degree of atrophy in secondary progressive versus relapsing-remitting patients, approaching statistical significance (see Results).

TABLE 2: Comparing two methods of measuring brain parenchymal fraction and their association with clinical and MR imaging variables in 52 patients with multiple sclerosis

	Semiautomated		Automated	
	2D	3D	2D	3D
Third ventricular width	$r = -.82$, $P < .001$	$r = -.79$, $P < .001$	NR	$r = -.81$, $P < .001$
Bicaudate ratio	$r = -.74$, $P < .001$	$r = -.74$, $P < .001$	NR	$r = -.74$, $P < .001$
T1 hypointense lesion volume	$r = -.38$, $P = .006$	$r = -.44$, $P = .001$	NR	$r = -.48$, $P < .001$
FLAIR hyperintense lesion volume	$r = -.24$, $P = .09$	$r = -.35$, $P = .01$	NR	$r = -.44$, $P = .001$
EDSS	$r = -.44$, $P < .001$	$r = -.47$, $P < .001$	NR	$r = -.33$, $P = .008$
Disease duration	$r = -.50$, $P < .001$	$r = -.39$, $P = .002$	NR	$r = -.53$, $P < .001$

techniques showed a high rate of reproducibility and appear to have a similar degree of validity as estimated by their relationship to established clinical and MR imaging disease markers (Tables 1 and 2; Figs 4 and 5). Both methods were able to detect whole-brain atrophy in MS versus healthy controls (Table 1; Fig 5). Both methods showed a similar degree of validity as estimated against linear measures of brain atrophy such as third ventricular width (Table 2; Fig 6) and bicaudate ratio (Table 2; Fig 7). Both methods showed a similar correlation with total brain T1 hypointense and FLAIR hyperintense lesion volumes (Table 2; Figs 8 and 9). The two atrophy methods, however, were not completely interchangeable in terms of correlation with other MR imaging and clinical measures (Tables 1 and 2; Figs 5–9). The slight differences are most likely due to random effects or differences in masking and segmentation steps. Such small variations are not likely to be of clinical relevance.

In terms of BPFs performed on two different pulse sequences (2D spin-echo and 3D gradient echo), semiautomated BPFs showed similar and interchangeable results in terms of the data obtained and correlation with MR imaging and clinical variables (Tables 1 and 2, Figs 4–9). In terms of BPFs performed on the same pulse sequence but different segmentation procedures, the semiautomated BPF generated data remarkably similar to the automated BPF including comparable estimates of validity by using the 3D image series (Tables 1 and 2; Figs 4–9).

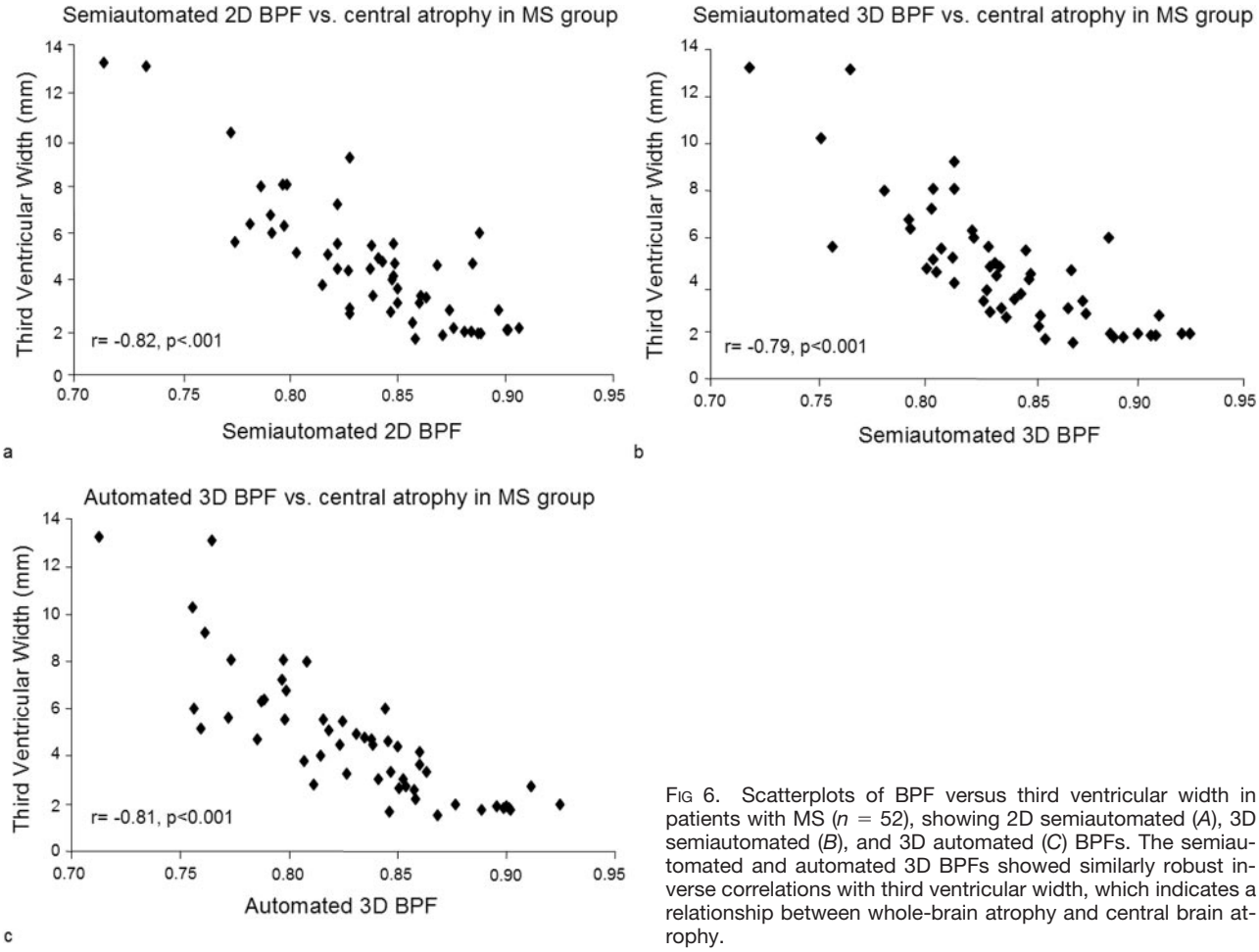

FIG 6. Scatterplots of BPF versus third ventricular width in patients with MS ($n = 52$), showing 2D semiautomated (*A*), 3D semiautomated (*B*), and 3D automated (*C*) BPFs. The semiautomated and automated 3D BPFs showed similarly robust inverse correlations with third ventricular width, which indicates a relationship between whole-brain atrophy and central brain atrophy.

Some comments are warranted regarding the effects of section thickness and partial volume effects. We chose to study 5-mm 2D and 2.5-mm 3D acquisitions in this study. The 5-mm sequence was chosen to replicate a fairly efficient protocol that is typically used in the routine clinical care of patients with MS (if imaging time permits, many centers use 3-mm 2D sequences instead). The 2.5-mm 3D sequence was chosen to determine the effect of higher resolution acquisition with an imaging time that was clinically feasible (many centers use thinner-section 3D source images for atrophy research studies). We expected that the 5-mm sections would introduce partial volume effects that would make them less useful for the assessment of BPF (28). This was true for the automated algorithm, which gave unreliable results because of poor segmentation. For the semiautomated method comparing 5-mm 2D and 2.5-mm 3D images, however, the BPFs were remarkably similar and interchangeable (Tables 1 and 2; Figs 4–9), which suggests that partial volume effects did not contribute significantly to the quality of segmentation. Nevertheless, if we would have optimized the 3D sequence with the trade-off of longer imaging time (eg, 1-mm thickness), we might have seen differences related to partial volume averaging (28). Another potential pitfall we addressed related to partial volume effects was section selection and coverage for the 5-mm 2D sequence. We chose a 24-section coverage scheme, which saved 2 minutes of imaging time per patient and was based on identifying the inferior border of the vermis on axial views as the first caudal section (Fig 1). This could have introduced partial volume errors. When we performed BPF on full-brain coverage sections (28 sections), however, we obtained similar BPFs to those obtained from the 24-section routine. This suggests the increased efficiency that can be obtained in the semiautomated approach, but it would be informative to perform further studies applying various segmentation software programs on the same images or the same software by using different images to extend our findings and provide additional data on sensitivity, validity, and reliability.

Previous studies showed the usefulness of both automated and semiautomated segmentation methods in measuring brain atrophy (1, 8, 9, 11–19, 22, 25, 26, 28, 29, 31, 32, 40–43). Both methods are widely available and simple to implement by using a variety of image analysis software packages. There are likely explanations for the trade-offs between the two approaches. In short, the semiautomated method has the potential advantage of accuracy, whereas the automated method has the advantage of efficiency. A semiautomated method allows user interaction to en-

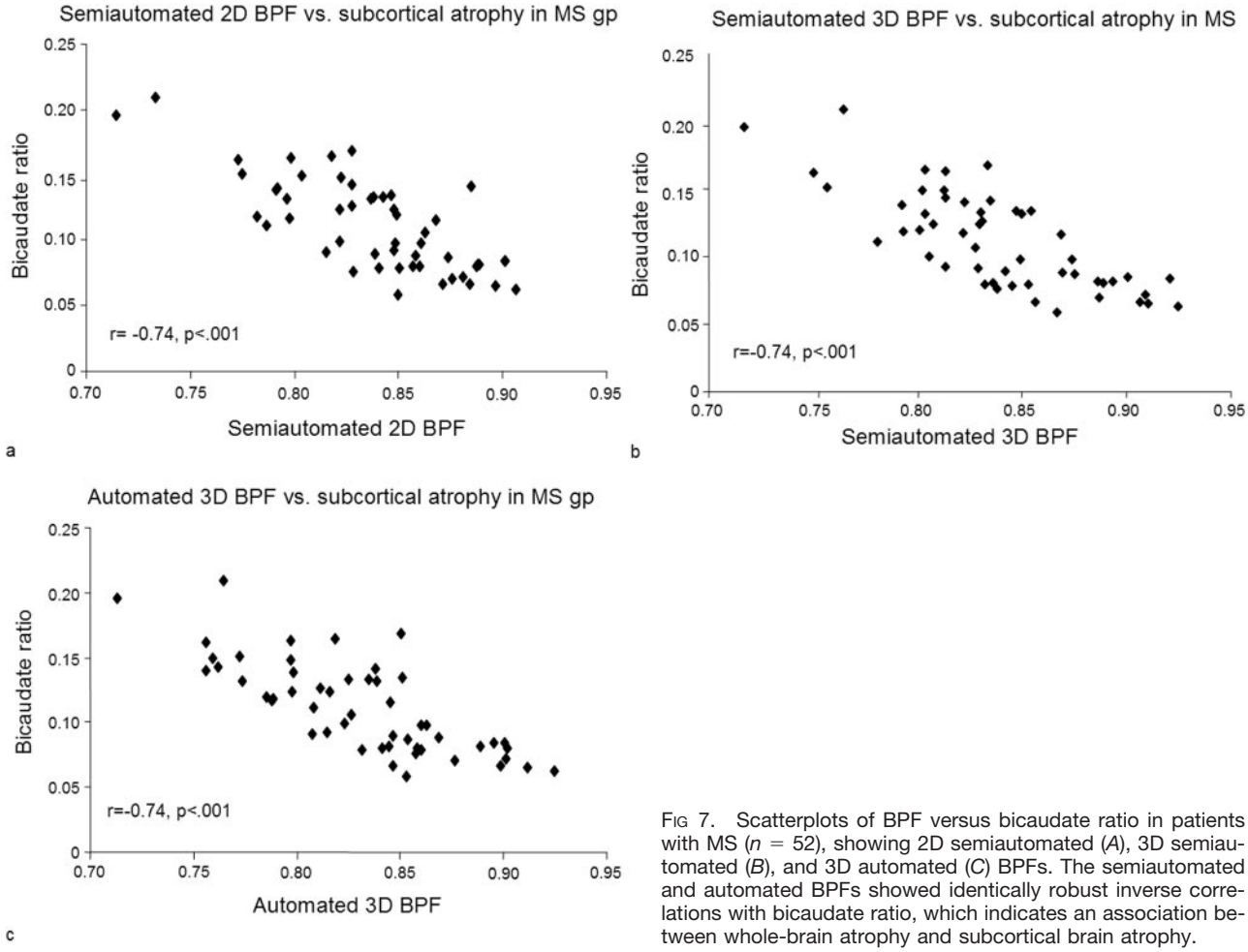

FIG 7. Scatterplots of BPF versus bicaudate ratio in patients with MS ($n = 52$), showing 2D semiautomated (*A*), 3D semiautomated (*B*), and 3D automated (*C*) BPFs. The semiautomated and automated BPFs showed identically robust inverse correlations with bicaudate ratio, which indicates an association between whole-brain atrophy and subcortical brain atrophy.

sure that masking and segmentation has been performed precisely (Figs 1 and 2). Examples of corrections that are typically made by the operator related to masking include reduction of undersampling (making sure that the brain parenchyma has been kept intact) and reduction of oversampling (making sure that the extracranial tissue has been removed). The operator interaction, however, introduces variability that reduces the reproducibility of the technique. In contrast, the automated method has a high degree of reliability and requires only minimal operator time (Fig 3). The automated method we tested also has the advantage of separating the brain parenchyma into gray matter and white matter compartments, allowing a compartmental analysis of gray versus white matter atrophy (1, 8, 30). Despite the inherent differences between these methods, there were remarkable similarities in comparison of the BPFs obtained, which showed a high degree of intercorrelation, similar differences versus normal controls and similar correlation with other disease measures (Tables 1 and 2, Figs 4–9). The slight differences are most likely due to random effects or slight differences in masking and segmentation steps. Regarding the segmentation of the intracranial volume into brain parenchyma versus CSF, the presence of MS lesions may lead to misclassification of brain parenchyma as CSF and each method may apply these misclassifications to a different extent (8, 30, 38); however, lesion misclassification did not seem to be a major factor in the present study by using the semiautomated or automated method.

There are several other factors that can potentially affect the sensitivity, validity, and reliability of whole-brain atrophy data, including type of pulse sequence (26, 32), section thickness, field strength (43), the patient's medical status (eg, nutrition, hydration, body habitus), and medications such as corticosteroids (17, 40). The advantage of using a semiautomated method is that the segmentation approach can be easily adjusted for different pulse sequences and for system upgrades. A preliminary study of 34 patients with relapsing-remitting MS compared two semiautomated and one automated method of determining BPF (31). Whereas the automated method used was different from that used in the present study, the semiautomated methods showed a significant correlation with physical disability and MR imaging lesion volumes whereas the automated method did not (31). Controls were not examined. In the present study, we have extended previous findings by examining relapsing-remitting versus secondary progressive groups and MS versus healthy volunteers groups. A comparison of various methods should be

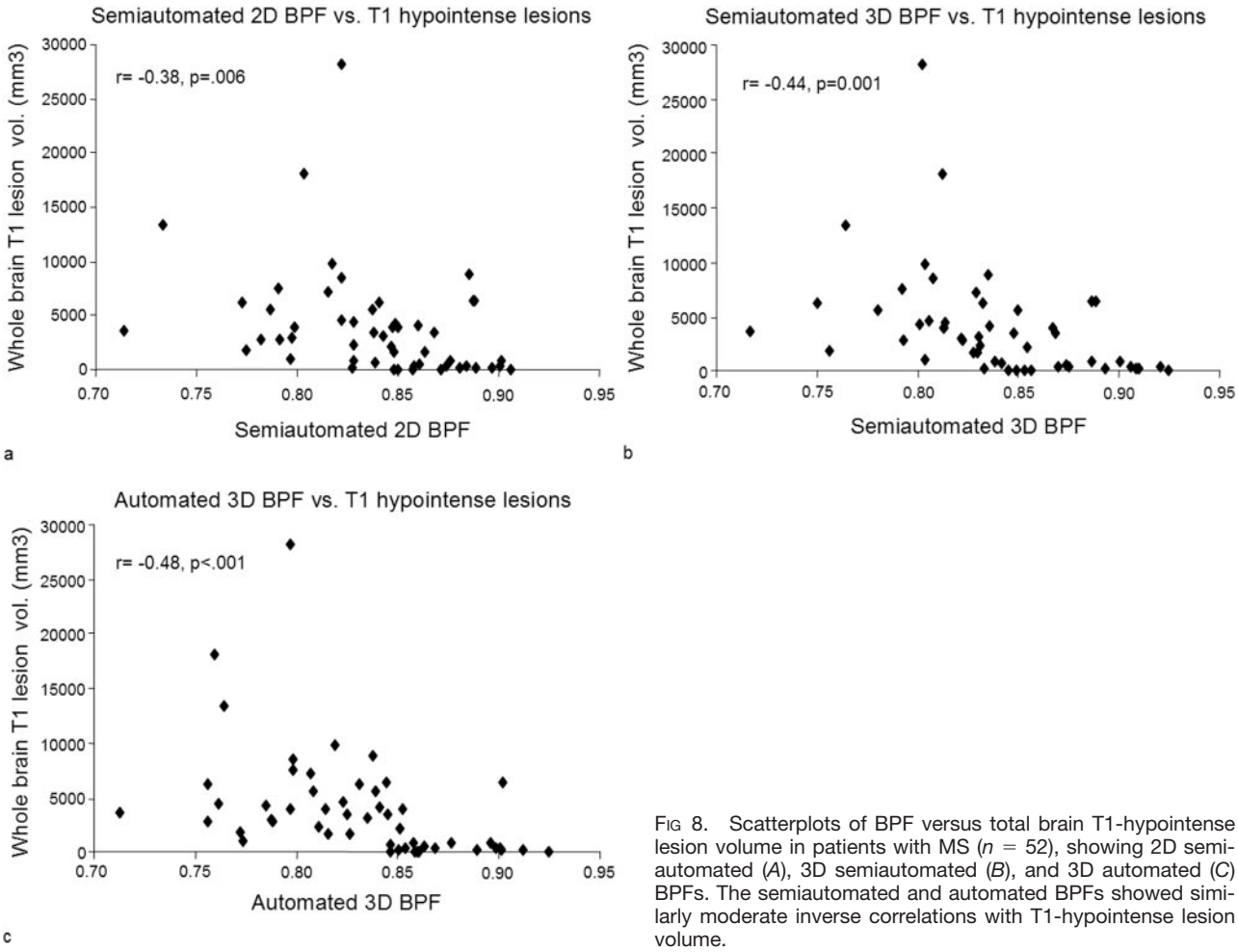

FIG 8. Scatterplots of BPF versus total brain T1-hypointense lesion volume in patients with MS ($n = 52$), showing 2D semiautomated (*A*), 3D semiautomated (*B*), and 3D automated (*C*) BPFs. The semiautomated and automated BPFs showed similarly moderate inverse correlations with T1-hypointense lesion volume.

performed in longitudinal studies to better characterize the sensitivity and clinical relevance of differences in these methods.

MS is now recognized as a globally destructive disease process involving tissue loss, tract degeneration, and atrophy of both the brain and spinal cord (1). Therefore, whole-brain atrophy is of growing interest as a reliable sensitive marker for assessing disease progression and to monitor clinical trials (1, 28). Brain atrophy is a common finding in patients with MS and has been shown to occur early in the disease course (1, 7–20, 28). Demyelination, axonal loss, reduced axonal density, neuronal loss, contraction from astrogliosis, apoptosis, iron deposition, and Wallerian degeneration are factors proposed to contribute to atrophy (1, 4–7, 14, 15, 17, 28, 37, 44). The development of atrophy seems to depend both on the extent of global damage and on focal (overt lesion-related) damage (1, 2, 14, 28, 29). Recent studies show that axonal loss is present in normal-appearing white and gray matter (4, 6). Wallerian degeneration secondary to axonal transection at the site of inflammation and in normal-appearing white matter most likely contributes to global tissue loss (1–4, 7, 28). T2 hypointensity in gray matter, suggestive of pathologic iron deposition, has also been related to brain atrophy in MS (5, 37, 44). The pathologic characteristics of MS are thus more complex than those of lesions depicted on conventional T1- and T2-weighted MR images. The most often used T2-hyperintense and gadolinium-enhancing lesion markers represent various stages of lesion formation but are nonspecific for a wide range of disease and show unreliable correlation with physical disability and other clinical findings (2, 20, 45–49). Brain atrophy has recently shown better clinical predictive value when compared with lesion measures (1, 18–25, 29). Thus, the detection of CNS atrophy and its quantification by MR imaging will likely continue to interest the MS scientific community.

Conclusion

Despite the presumption that automated measures of whole-brain atrophy are preferable to semiautomated methods, the two provided similar and nearly interchangeable data. The similarities between the two methods persist after controlling for the pulse sequence employed to generate images. Overall, these measures behave similarly in discriminating MS from controls, both showing a high level of reproducibility, strong correlation with linear atrophy measures, and a moderate level of correlation with disability and lesion measures. The potential advantage of using the semiautomated method is its accuracy

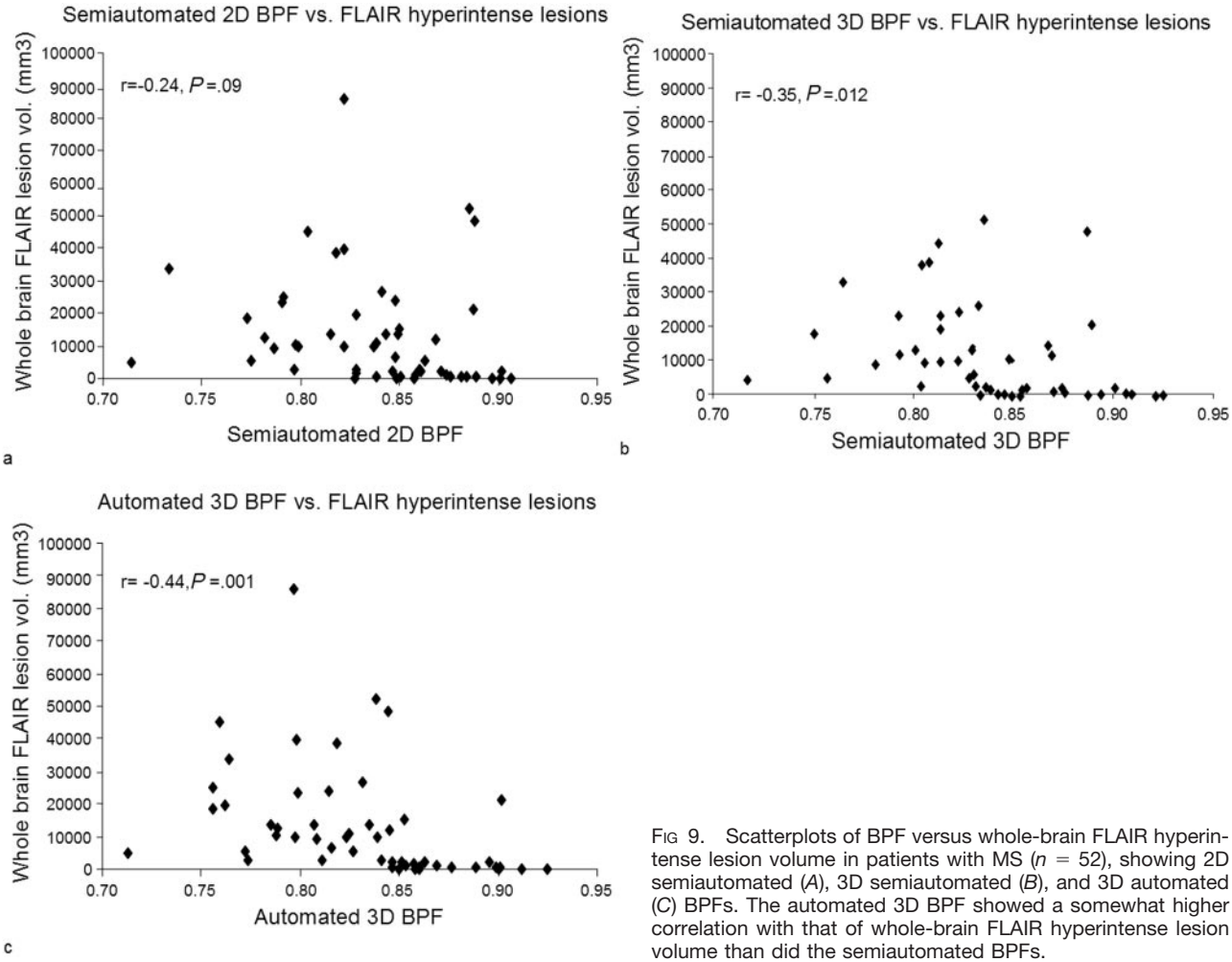

FIG 9. Scatterplots of BPF versus whole-brain FLAIR hyperintense lesion volume in patients with MS ($n = 52$), showing 2D semiautomated (*A*), 3D semiautomated (*B*), and 3D automated (*C*) BPFs. The automated 3D BPF showed a somewhat higher correlation with that of whole-brain FLAIR hyperintense lesion volume than did the semiautomated BPFs.

and precision and that it is easily adjustable, whereas the automated method is more efficient and reproducible. Longitudinal studies are warranted to further compare the validity, sensitivity, and reliability of these and related methods.

Acknowledgments

Our work was supported by research grants from the National Institutes of Health (NIH-NINDS grant K23 NS42379–01, to R.B.), the National Multiple Sclerosis Society (RG 3258A2/1, to B.W.-G. and R.B.), the National Science Foundation (DBI-0234895, to B.W.-G. and R.B.) and the University at Buffalo, State University of New York, Medical Student Research Fellowship (M.S.). We thank Robert A. Bermel, Christopher W. Tjoa, and Jin Kuwata for technician support, and Mark Horsfield, PhD, for developing and installing the neuroimaging analysis software.

References

1. Zivadinov R, Bakshi R. **Role of MRI in multiple sclerosis. II. Brain and spinal cord atrophy.** *Front Biosci* 2004;9:647–664
2. Zivadinov R, Bakshi R. **Role of MRI in multiple sclerosis. I. inflammation and lesions**. *Front Biosci* 2004;9:665–683
3. Bakshi R, Miletich RS, Kinkel PR, et al. **High-resolution fluorodeoxyglucose positron emission tomography shows both global and regional cerebral hypometabolism in multiple sclerosis.** *J Neuroimaging* 1998;8:228–234
4. Cifelli A, Arridge M, Jezzard P, et al. **Thalamic neurodegeneration in multiple sclerosis.** *Ann Neurol* 2002;52:650–653
5. Bakshi R, Dmochowski J, Shaikh ZA, Jacobs L. **Gray matter T2 hypointensity is related to plaques and atrophy in brains of multiple sclerosis patients**. *J Neurol Sci* 2001;185:19–26
6. Trapp BD, Peterson J, Ransohoff R, et al. **Axonal transection in the lesions of multiple sclerosis.** *N Engl J Med* 1998;338:278–285
7. Simon JH, Kinkel RP, Jacobs L, et al. **A Wallerian degeneration pattern in patients at risk for MS.** *Neurology* 2000;5:1155–1160
8. Chard D, Griffin CM, Parker GJ, et al. **Brain atrophy in clinically early relapsing-remitting multiple sclerosis.** *Brain* 2002;125:327–337
9. Brex PA, Jenkins R, Fox NC, et al. **Detection of ventricular enlargement in patients at the earliest clinical stage of MS.** *Neurology* 2000;54:1689–1691
10. Brex PA, Leary SM, O'Riordan JI, et al. **Measurement of spinal cord area in clinically isolated syndromes suggestive of multiple sclerosis.** *J Neurol Neurosurg Psychiatry* 2001;70:544–547
11. Losseff NA, Wang L, Lai HM, et al. **Progressive cerebral atrophy in multiple sclerosis: a serial MRI study.** *Brain* 1996;119:2009–2019
12. Rudick RA, Fisher E, Lee JC, et al. **Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS**. *Neurology* 1999;53:1698–704
13. Ge Y, Grossman RI, Udupa JK, et al. **Brain atrophy in relapsing-remitting multiple sclerosis and secondary progressive multiple sclerosis: longitudinal quantitative analysis.** *Radiology* 2000;214:665–70
14. Simon JH, Jacobs LD, Campion MK, et al. **A longitudinal study of brain atrophy in relapsing multiple sclerosis.** *Neurology* 1999;53:139–148
15. Rovaris M, Comi G, Rocca MA, et al. **Short-term brain volume change in relapsing-remitting multiple sclerosis: effect of glatiramer acetate and implications.** *Brain* 2001;124:1803–1812

16. Filippi M, Rovaris M, Iannucci G, et al. **Whole brain volume changes in patients with progressive MS treated with cladribine.** *Neurology* 2000;55:1714–1717
17. Zivadinov R, Rudick RA, De Masi R, et al. **Effects of IV methylprednisolone on brain atrophy in relapsing-remitting MS.** *Neurology* 2001;57:1239–1247
18. Zivadinov R, Sepcic J, Nasuelli D, et al. **A longitudinal study of brain atrophy and cognitive disturbances in the early phase of relapsing-remitting multiple sclerosis.** *J Neurol Neurosurg Psychiatry* 2001;70:773–780
19. Fisher E, Rudick RA, Simon JH, et al. **Eight-year follow-up study of brain atrophy in patients with MS.** *Neurology* 2002;59:1412–1420
20. Bakshi R, Benedict RH, Bermel RA, Jacobs L. **Regional brain atrophy is associated with physical disability in multiple sclerosis: semiquantitative magnetic resonance imaging and clinical findings.** *J Neuroimaging* 2001;11:129–136
21. Benedict RHB, Weinstock-Guttman B, Fishman I, et al. **Prediction of neuropsychological impairment in multiple sclerosis: comparison of conventional magnetic resonance imaging measures of atrophy and lesion burden.** *Arch Neurol* 2004;61:226–230
22. Bermel R, Bakshi R, Tjoa C, et al. **Bicaudate ratio as an MRI marker of brain atrophy in multiple sclerosis.** *Arch Neurol* 2002; 59:275–280
23. Janardhan V, Bakshi R. **Quality of life and its relationship to brain lesions and atrophy on magnetic resonance images in 60 patients with multiple sclerosis.** *Arch Neurol* 2000;57:1485–1491
24. Bakshi R, Czarnecki D, Shaikh ZA, et al. **Brain MRI lesions and atrophy are related to depression in multiple sclerosis.** *Neuroreport* 2000;11:1153–1158
25. Bermel RA, Sharma J, Tjoa CW, et al. **A semiautomated measure of whole-brain atrophy in multiple sclerosis.** *J Neurol Sci* 2003;208: 57–65
26. Leigh R, Ostuni J, Pham D, et al. **Estimating cerebral atrophy in multiple sclerosis patients from various MR pulse sequences.** *Mult Scler* 2002;8:420–429
27. Zorzon M, de Masi R, Nasuelli D, et al. **Depression and anxiety in multiple sclerosis: a clinical and MRI study in 95 subjects.** *J Neurol* 2001;248:416–421
28. Miller DH, Barkhof F, Frank JA, et al. **Measurement of atrophy in multiple sclerosis: pathological basis, methodological aspects and clinical relevance.** *Brain* 2002;125:1676–1695
29. Kalkers NF, Vrenken H, Uitdehaag BMJ, et al. **Brain atrophy in multiple sclerosis: impact of lesions and damage of whole brain tissue**. *Mult Scler* 2002;8:410–414
30. Chard DT, Parker GJ, Griffin CM, et al. **The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology.** *J Magn Reson Imaging* 2002;15:259–267
31. Zivadinov R, Bakshi R, Grop A, et al. **A comparison of different quantitative MRI techniques to measure whole brain atrophy in relapsing-remitting multiple sclerosis.** *Mult Scler* 2002;8:s99
32. Horsfield MA, Rovaris M, Rocca MA, et al. **Whole brain atrophy in multiple sclerosis measured by two segmentation processes from various MRI sequences.** *J Neurol Sci* 2003;216:169–177
33. Kurtzke JF. **Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS).** *Neurology* 1983;33:1444–1452
34. Bakshi R, Caruthers SD, Janardhan V, Wasay M. **Intraventricular CSF pulsation artifact on fast fluid-attenuated inversion-recovery MR images: analysis of 100 consecutive normal studies.** *AJNR Am J Neuroradiol* 2000;21:503–508
35. Bakshi R, Ariyaratana S, Benedict RHB, Jacobs L. **Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions.** *Arch Neurol* 2001; 58:742–748
36. 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
37. Bakshi R, Benedict RHB, Bermel RA, et al. **T2 hypointensity in the deep gray matter of patients with multiple sclerosis: a quantitative magnetic resonance imaging study.** *Arch Neurol* 2002;59:62–68
38. Sanfilipo MP, Benedict RHB, Zivadinov R, Bakshi R. **Correction for intracranial volume in analysis of whole brain volume in multiple sclerosis: the proportion vs. residual method.** *Neuroimage* (in press)
39. Ashburner J, Friston K. **Voxel-based morphometry: the methods**. *Neuroimage* 2000;11:805–821
40. Hoogervorst EL, Polman CH, Barkhof F. **Cerebral volume changes in multiple sclerosis patients treated with high-dose intravenous methylprednisolone.** *Mult Scler* 2002;8:415–419
41. Hohol MJ, Guttmann CR, Orav J, et al. **Serial neuropsychological assessment and magnetic resonance imaging analysis in multiple sclerosis.** *Arch Neurol* 1997;54:1018–1025
42. Pelletier D, Nelson SJ, Oh J, et al. **MRI lesion volume heterogeneity in primary progressive MS in relation with axonal damage and brain atrophy.** *J Neurol Neurosurg Psychiatry* 2003;74:950–952
43. Gasperini C, Rovaris M, Sormani MP, et al. **Intra-observer, inter-observer and inter-scanner variations in brain MRI volume measurements in multiple sclerosis.** *Mult Scler* 2001;7:27–31
44. Bakshi R, Puli SR. Tjoa CW, et al. **Gray matter T2 hypointensity in multiple sclerosis: a 2-year longitudinal clinical–MRI study of 79 patients.** *Neurology* 2002;58:A154
45. Simon JH. **Contrast-enhanced MR imaging in the evaluation of treatment response and prediction of outcome in multiple sclerosis.** *J Magn Reson Imaging* 1997;7:29–37
46. Katz D. Taubenberger JK, Cannella B, et al. **Correlation between magnetic resonance imaging findings and lesion development in chronic, active multiple sclerosis.** *Ann Neurol* 1993;34:661–669
47. Barnes D, Munro PM, Youl BD, et al. **The longstanding MS lesion: a quantitative MRI and electron microscopic study.** *Brain* 1991; 114:1271–1280
48. Barkhof F. **MRI in multiple sclerosis: correlation with expanded disability status scale (EDSS).** *Mult Scler* 1999;5:283–286
49. Bakshi R, Miletich RS, Henschel K, et al. **Fatigue in multiple sclerosis: cross sectional correlation with brain MRI findings in 71 patients.** *Neurology* 1999;53:1151–1153