MR Volume Segmentation of Gray Matter and White Matter Using Manual Thresholding: Dependence on Image Brightness

Gordon J. Harris, Patrick E. Barta, Luon W. Peng, Seong Lee, Paul D. Brettschneider, Amish Shah, Jeffery D. Henderer, Thomas E. Schlaepfer, and Godfrey D. Pearlson

PURPOSE: To describe a quantitative MR imaging segmentation method for determination of the volume of cerebrospinal fluid, gray matter, and white matter in living human brain, and to determine the method's reliability. **METHODS:** We developed a computer method that allows rapid, user-friendly determination of cerebrospinal fluid, gray matter, and white matter volumes in a reliable manner, both globally and regionally. This method was applied to a large control population (N = 57). **RESULTS:** Initially, image brightness had a strong correlation with the gray-white ratio ($r = .78$). Bright images tended to overestimate, dim images to underestimate gray matter volumes. This artifact was corrected for by offsetting each image to an approximately equal brightness. After brightness correction, gray-white ratio was correlated with age ($r = -.35$). The age-dependent gray-white ratio was similar to that for the same age range in a prior neuropathology report. Interrater reliability was high (.93 intraclass correlation coefficient). **CONCLUSIONS:** The method described here for gray matter, white matter, and cerebrospinal fluid volume calculation is reliable and valid. A correction method for an artifact related to image brightness was developed.

Index terms: Magnetic resonance, experimental; Magnetic resonance, technique; Magnetic resonance, tissue characterization; Brain, magnetic resonance; Brain, volume; Cerebrospinal fluid; Gray matter; White matter

AJNR Am J Neuroradiol 15:225–230, Feb 1994

The purpose of this study was to present a method for quantitative segmentation of cerebrospinal fluid (CSF), gray matter, and white matter volumes from live human subjects using magnetic resonance (MR) images and to determine the method's reliability. In a previous report, Harris et al described a rapid, user-friendly method for reliable determination of brain and CSF volumes from transaxial MR images (1). The current report expands on this previous work, which used T2 minus proton-weighted MR images to distinguish CSF from the brain. The current method first corrects the MR images for radio frequency field inhomogeneity after extracerebral regions and CSF have been stripped from the images, then presents T2 plus proton-weighted images of brain alone, without CSF or background, for gray-white matter segmentation by interactive thresholding.

In the current report, we applied gray-white segmentation to the study of a large control population and found an artifact related to manual gray-white threshold segmentation. The gray-white ratio was strongly correlated with the mean gray-white threshold setting, which is determined by the brightness of the displayed image. We suggest a method for correction of this artifact. The dependence of gray-white ratio on age was assessed both before and after image brightness correction. The age-dependent gray-white ratio determined in this study was compared with that reported in the neuropathology literature (2).

Received November 20, 1992; accepted pending revision January 19, 1993; revision received March 3.

From the Division of Psychiatric Neuroimaging, The Department of Psychiatry and Behavioral Sciences, The Johns Hopkins Medical Institutions, Baltimore, Md; and Neuroimaging Research Laboratory, Department of Psychiatry, Tufts University School of Medicine, Boston, Mass (G.J.H.).

Address reprint requests Gordon J. Harris, Ph.D., Director, Neuroimaging Research Laboratory, Department of Psychiatry, Tufts University School of Medicine, Boston, MA 02111.

AJNR 15:225–230, Feb 1994 0195-6108/94/1502–0225

Methods

Subjects

The subject group consisted of 57 persons (42 men and 15 women), age 31.5 ± 7.9 years (range 19 to 47), who had no history of current or past substance abuse or neurologic illness, including head trauma causing unconsciousness for longer than 1 hour. No subject had a family history of major psychiatric illness. Exclusion criteria also included personal history of any psychiatric illness as assessed using the structured clinical interview for the *Diagnostic and Statistical Manual of Mental Disorders*, third edition, revised (3). This project was approved by the Johns Hopkins Joint Committee on Clinical Investigations, and informed consent was obtained from all subjects.

Equipment and Scan Protocol

The MR images used were acquired on a General Electric (Milwaukee, Wis) 1.5-T Signa scanner and transported on nine-track magnetic tape to our computer system. The analysis software described in this paper was developed in house in C and XWindows on a Digital Equipment (Nashua, NH) DECstation 3100 work station with a color graphics monitor. The work station is a client of our DECsystem 5500 server, which operates Ultrix-32 version 4.1. The studies were archived on read/write magnetooptical disks (300 megabytes per side).

Contiguous, 5-mm-thick axial sections were acquired with simultaneous T2 and proton sequences. The images extended from the base of the cerebellum to the vertex, parallel to the anterior commissure–posterior commissure line. Scan parameters were 2500/20/1 (repetition time/echo time/excitations) for proton and 2500/80/1 for T2 images.

CSF-Brain Segmentation

A brief description of the CSF segmentation is presented here. The method, its reliability, and its validity in phantom studies have been presented in detail previously (1). In T2-weighted images, CSF appears bright compared with brain; in proton-weighted images, CSF appears dark. Therefore, in T2 minus proton-weighted images, CSF is preferentially highlighted compared with brain. T2-weighted images have lower pixel values than proton-weighted images, however, so a constant must be added to the different images to create positive pixel values for display. In T2 minus proton-weighted images with a constant added, the background appears bright, which is not desired. Therefore, all pixels below a background threshold in the T2 image were left as T2 level background; all other pixels (includes brain, CSF, and scalp) used T2 minus proton weighting. In this manner the boundary between skull and brain/CSF remained clear. This is a subtle but important point.

For brain and CSF volume calculation, the rater began at the bottom-most T2 minus proton-weighted image and worked toward the top, image by image. All intracranial regions, including all CSF, were selected by a threshold-guided, semiautomatic edge-follower algorithm (1). In this manner, quick, accurate, and reliable calculations of total brain tissue and CSF volumes were done for the entire intracranial space. All extracerebral pixels were then set to zero to identify them as background. The CSF was then highlighted, and pixels containing CSF were set to a value of one to flag them to a given nonbrain value (brain pixels were generally in the range of 20 to 40 in our 64 gray level display). These images, with CSF set to 1 and background set to zero, were then saved and sent to the next step in the process the radio frequency inhomogeneity correction filter.

Radio Frequency Inhomogeneity Correction

Radio frequency inhomogeneity introduces a low-frequency gradient across the image, which makes thresholding spatially dependent. Lim and Pfefferbaum (4) corrected this by low-pass filtering the original image, then subtracting the low-pass filter from the original image. High-contrast edges in the image produce a ringing artifact in the filtered image, necessitating background identification. The background must be set so that edge artifacts are not introduced. Therefore, we first removed background and CSF from the image. Before radio frequency correction filtering, all pixels representing background or CSF were set to the mean value of the brain, to eliminate edges that cause ringing artifacts around the brain edge or ventricles during radio frequency correction filtering. Our method has the advantage over that of Lim and Pfefferbaum (4) of avoiding ringing around both the ventricles and the outer brain edge. After radio frequency inhomogeneity correction, background was set back to 0 and CSF back to 1 before the final step, gray-white segmentation. The effects of radio frequency inhomogeneity on segmentation can be seen in Figure 1.

Gray-White Matter Segmentation

As described by Lim and Pfefferbaum (4), gray-white matter contrast is improved if T2 plus proton-weighted images are used. After radio frequency inhomogeneity correction, it becomes possible to threshold such summed images for gray and white matter. The summed image values were first divided by a constant scaling factor to decrease the pixel intensity range, which was too large for the number of gray levels used. A subtraction constant was then used to reduce the image brightness to place the intensity within the constraints of the gray scale for image display (in this case, 64 levels). If only a scaling factor were used, it would necessarily be large and would compress the dynamic range of the image values into a small set of intensity levels, making threshold setting difficult. Hence, a smaller scaling factor combined with a subtraction factor was preferable. It caused less compression of the dynamic range of the summed image, while keeping the image within the gray scale range. Selection of the subtraction factor is further discussed in the next subsection.

Fig. 1. *Left,* Three-color representation of gray matter, white matter, and CSF segmentation without radio frequency inhomogeneity correction. Note the diagonal gradient, which makes part of the image too dark and part too light (from Pearlson and Marsh [12]).

Center, Three-color representation of the same segmented image after radio frequency inhomogeneity correction. Note the improvement in spatial uniformity of gray and white matter thresholding (from Pearlson and Marsh [12]).

Right, Lines defining the intracranial boundary and cortical areas with five subregions per hemisphere, and the inner section of subcortical gray and ventricles for regional volume calculations.

Our gray-white segmentation program used interactive threshold determination for each section. A starting threshold level was provided, which could be incremented and decremented by the rater. The gray matter pixels were highlighted in color based on the threshold. The color could be toggled on and off to view the image. When the desired threshold was selected, a three-color mask could be viewed of the image, identifying gray, white, and CSF pixels, as shown in Fig 1. This mask could be toggled on or off, and the threshold again could be manipulated until the optimal threshold was reached, as judged by the rater based on visual correspondence between the gray matter in the image and those portions highlighted in color. An output data file was created, which contained study name, field of view, and section thickness, the areas in square millimeters of gray matter, white matter, brain (gray matter plus white matter), CSF, and total (brain plus CSF), the threshold chosen for each section, and the volume totals (total area multiplied by section thickness). Percent CSF, gray matter, and white matter were calculated, as was the gray-white ratio.

Correction for Image Brightness

Initially, we used the same subtraction factor for all sections of all images to set brightness such that the gray-white threshold was in the range of a gray value of 40 (0 is black; 63 is white). This range provides the best visual contrast between gray matter and white matter. Most of the images had a mean gray-white threshold near 40. However there were some images less than 30 and some greater than 50 (mean gray-white threshold was calculated by averaging the gray-white threshold for each section). It was noted that the gray-white ratios were unusually low for those images with low gray-white threshold settings and unusually high for those images with high settings. After plotting mean gray-white threshold with gray-white ratio for a large set of control subjects, it became clear that there was a problem with this approach, that the gray-white ratio was related to image brightness, as seen in Figure 2. We also noted variability of image brightness between sections within an image set. In order to correct this bias, the mean brain value (includes gray and white matter, excludes CSF and extracerebral regions) of each section was calculated and set to a gray value of approximately 40 by adjusting the image display subtraction factor. The images were resegmented, and the artifact disappeared as shown in Figure 2. This correction also made segmentation faster because there was very little variation of the threshold setting between sections and between individuals.

Regional Segmentation

In order to calculate the regional volumes of gray matter, white matter, and CSF, we defined an outer ring, which

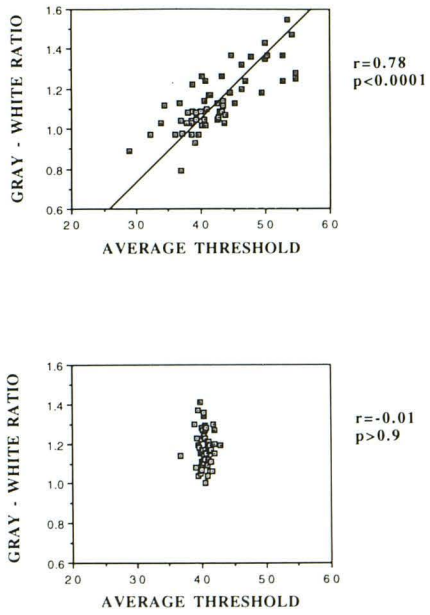

Fig. 2. The *top graph* illustrates the relationship between gray-white ratio and threshold value before brightness correction (the regression equation line has been drawn on this graph). This correlation was strongly significant ($r = .78$; $P < .0001$). The *bottom graph* shows the same variables plotted after brightness correction. The ranges of threshold values and of gray-white ratio values have been reduced, and the dependence of gray-white ratio on image brightness has been eliminated.

included the cortical gray matter and sulcal CSF, and an inner region, which contained subcortical gray matter and the ventricles. For each section on which regional segmentation was performed, the frontal and occipital poles were first defined by the rater using a mouse-driven cursor to define the interhemispheric fissure. The angle of this line was calculated and used as an offset angle for the regional cortical measures so that regions were evenly defined on each hemisphere. The midpoint of this line was used as the center point of the section.

The outer region was then defined by searching inward radially from extracranial space toward the center of the brain section every 6° to find the outer edge of the CSF and the outer edge of brain. An inner boundary was next defined which extended 2 cm in from the midpoint between the edge of the brain and the outer limit of CSF. The outer boundary was the intracranial edge. The cortical ring was divided into five equiangular regions per hemisphere (using the offset angle defined by the interhemispheric fissure line) in order to correspond to the regions defined for cortical circumferential profiling of emission computed tomography (5). Five regions per hemisphere were selected for our cortical segmentation and for the cortical circumferential profile method based on assessment of approximate Brodmann cortical functional areas as described previously (5). Other groups have similarly defined cortical regions based on four regions per hemisphere (6). The divisions of the brain on a selected section are shown in the *right* portion of Figure 1. No data using regional segmentation are presented in this article. It is presented as an option that will be used in future studies.

Reliability and Validity

The reliability and validity of CSF-brain segmentation has been reported previously (1). Interrater reliability of brain and CSF volumes had intraclass correlation coefficients of greater than 0.96, and validity was assessed using realistic brain phantom images. Therefore, interrater and intrarater reliability studies were done in this report for gray-white segmentation only (after correction for image brightness). Percent gray was used to calculate reliability for gray-white segmentation, because gray and white volumes are interdependent. For interrater reliability, two raters each blindly segmented 10 scans; for intrarater reliability, one rater blindly segmented 10 scans on two occasions each. Intraclass correlation coefficients were calculated for percent gray matter (gray matter volume · 100/ total intracranial volume). Validity was assessed by comparison of age-dependent gray-white ratios with those reported in a prior neuropathology study (2). Because of the convoluted nature of gray and white matter, it is difficult to assess validity accurately with a realistic phantom.

Results

Interrater reliability of gray matter volume calculation had an intraclass correlation coefficient of 0.93 based on gray percentages of 10 subjects each rated by two raters (after image brightness correction). Intrarater reliability intraclass correlation coefficient was .99.

There was a strong positive correlation (r = .78; $P < .0001$) between gray-white ratio and mean threshold before correction for image brightness. This correlation disappeared after brightness correction ($r = -.01$; not significant). The brightness effect can be clearly seen in Figure 2, where the top graph displays gray-white ratio values without correction, and the bottom graph is after brightness correction. Mean thresholds ranged from 28.96 to 54.79 before brightness correction, and from 36.66 to 42.85 after correction. Likewise, the gray-white ratio ranged from 0.79 to 1.55 before brightness correction and from 1.00 to 1.41 after. Thus, both the ranges of the thresholds and of the gray-white ratios were reduced by brightness correction.

The subject group was divided into thirds using uncorrected threshold values to examine what effect the brightness correction had on the gray-white ratio in the lower and upper third. The assumption was that the brightness correction would rectify the underassessment of gray volume in the lower third of thresholds and the overassessment of gray volume in the top third. The lower third had a mean gray-white ratio of 1.02 ± 0.10 before correction and 1.14 ± 0.09 after ($t = 4.56$; $P < .0001$). The upper third had a gray-white ratio of 1.28 ± 0.14 before correction and 1.21 ± 0.09 after ($t = 2.41$; $P < .03$). Therefore, these results are in agreement with our expectation. (All following analyses used postcorrection values unless specifically stated otherwise.)

Total mean intracranial volumes were 1434.6 ± 148.6 mL. Intracranial volume consisted of 50.5 ± 2.2% gray matter, 43.0 ± 2.1% white matter, and 6.5 ± 2.6% CSF. The average gray-white ratio for all 57 subjects after correction was 1.18 ± 0.09.

Regression analysis was performed to examine the relationship between gray-white ratio and age. Before brightness correction, no relationship was seen ($r = -.10$, not significant). There was a negative correlation between gray-white ratio and age after the images were brightness corrected ($r = -.35$; $P < .008$) indicating a decrease in gray matter relative to white matter with aging between the ages of 19 and 47, as shown in Figure 3. As a measure of validity, we compared the results of this study with those of a prior quantitative neuropathology report which used an image analyzer to determine gray matter and white

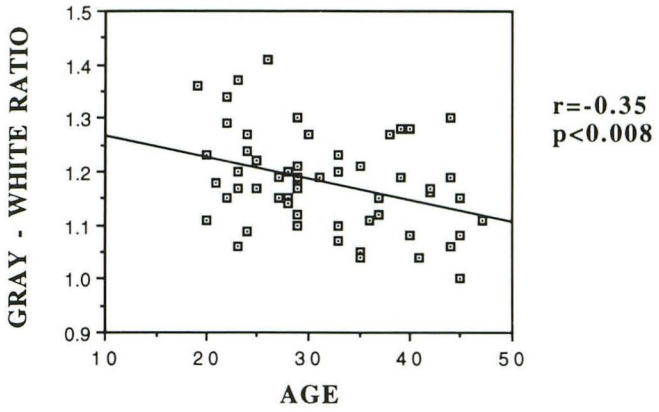

Fig. 3. Regression plot of gray-white ratio versus age after image brightness correction. There was significant negative correlation ($r = -.35$; $P < .02$). The regression line has been drawn on the image.

matter volumes at autopsy (2). In a group of 22 control subjects between the ages of 20 and 50, Miller and associates found that gray-white ratio decreased from about 1.3 at age 20 to about 1.1 at age 50. Based on the regression equation in our study, we estimate a gray-white ratio of 1.22 at age 20 decreasing to 1.11 at age 50. Our results are close to the actual neuropathology measurements described by Miller and associates.

Gray matter percentage had a strong negative correlation with age ($r = -.47$; $P < .0002$); white matter percentage had a slight but not significant increase with age ($r = .11$; not significant). CSF percentage increased significantly with age ($r = .32$; $P < .02$). Based on the regression equation, the expected value for gray percentage decreased from 52.0% to 48.0% between ages 20 and 50, whereas CSF percentage increased from 5.3% to 8.5% during this period. White matter increased slightly, from 42.6% to 43.5%.

Discussion

The method presented provides an easy-to-use, straightforward calculation of gray matter, white matter, and CSF fluid volumes using MR brain images. Our method has very good interrater and intrarater reliability and seems valid based on comparison with prior quantitative neuropathologic data (2). It was not possible to estimate directly the true accuracy of our volume measurements because our subject group is still living. Although previous validation of brain volume measurements using a brain-like phantom have been done (1), it is difficult to assess gray and white matter volumes accurately using a phantom, because of the convoluted morphology of gray and white matter.

The use of combined T2 and proton MR images presented in this study allows highlighting of different tissues by changing the relative combination of image types, as described above. These methods and scan parameters may not be optimal for gray/white matter differentiation, however, and other scan parameters which better distinguish gray/white differences might further improve accuracy.

Several groups have reported MR segmentation techniques based on combined T2-weighted and proton-weighted images. These methods use either manual thresholding (4), semiautomated thresholding based on the average of seed regions from two tissue types such as gray matter and white matter or CSF and brain (7), histogram-based best-fit semiautomated methods based on discriminant function analysis of seed regions (8), or semiautomated thresholding based on lookup values (9). One group has reported segmentation based on edge detection between tissue types, rather than thresholding (10, 11).

A decreased gray-white ratio associated with aging was also noted in prior MR segmentation reports (7, 8). These relative decreases in gray matter volume with aging are also associated with increased CSF volume (1, 7, 8).

We identified an artifact related to the manual thresholding of gray and white matter. For images that had low brightness, raters tended to define less gray matter; for bright images more gray matter was defined. This artifact resulted in a significant correlation between gray-white ratio and average threshold level. To avoid this bias artifact, we set all images to an equal intensity level to correct for brightness differences. This correction not only removed the artifact, but also made the program faster to operate. The brightness correction increased the low gray-white ratio found on dim images and decreased the elevated gray-white ratio on bright images. The fact that the age dependence of gray-white ratio became apparent after image brightness correction, and that the age-dependent values for gray-white ratio closely resembled those from a quantitative neuropathology report, supports the validity of this method using the brightness correction.

References

1. Harris GJ, Rhew EH, Noga T, Pearlson GD. User-friendly method for rapid brain and CSF volume calculation using transaxial MRI images. *Psychiaty Res: Neuroimag* 1991;40:61–68
2. Miller AKH, Alston RL, Corsellis JAN. Variation with age in the volumes of grey and white matter in the cerebral hemispheres of man: measurements with and image analyzer. *Neuropathol Appl Neurobiol* 1980;6:119–132

3. Spitzer RL, Williams JBW, Gibbon N, First MB. *Structured clinical interview for DSM-IIIR.* New York State Psychiatric Institute, Biometrics Research Department, 1989
4. Lim KO, Pfefferbaum A. Segmentation of MR brain images into cerebrospinal fluid spaces, white and gray matter. *J Comput Assist Tomogr* 1989;13:588–593
5. Harris GJ, Links JM, Pearlson GD, Camargo EE. Cortical circumferential profile of SPECT cerebral perfusion in Alzheimer's disease. *Psychiatry Res: Neuroimag* 1991;40:167–180
6. Zipursky RB, Lim KO, Sullivan EV, Brown BW, Pfefferbaum A. Widespread cerebral gray matter volume deficits in schizophrenia. *Arch Gen Psychiatry* 1992;49:195–205
7. Jernigan TL, Press GA, Hesselink JR. Methods for measuring brain morphologic features on magnetic resonance images: validation and normal aging. *Arch Neurol* 1990;47:27–32
8. Cohen G, Andreasen NC, Alliger R, et al. Segmentation techniques for the classification of brain tissue using magnetic resonance imaging. *Psychiatry Res: Neuroimag* 1992;45:33–51
9. Shenton ME, Kikinis R, McCarley RW, Metcalf D, Tieman J, Jolesz FA. Application of automated MRI volumetric measurement techniques to the ventricular system in schizophrenics and normal controls. *Schizophr Res* 1991;5:103–113
10. Filipek PA, Kennedy DN, Caviness VS, Rossnick SL, Spraggins TA, Starewicz PM. Magnetic resonance imaging-based brain morphometry: development and application to normal subjects. *Ann Neurol* 1989;25:61–67
11. Kennedy DN, Filipek PA, Caviness VS. Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Trans Med Imaging* 1989;8:1–7
12. Pearlson GD, Marsh L. Magnetic resonance imaging in psychiatry. In: Oldham JM, Riba MB, Tasman A, eds. *American psychiatric press review of psychiatry.* Vol 12. Washington, DC: American Psychiatric Press (in press)