Automated estimation of brain volume in Multiple Sclerosis with BICCR

Louis D. Collins, Johan Montagnat, A.P. Zijdenbos, Alan Evans

Automated estimation of brain volume in Multiple Sclerosis with BICCR

D. Louis Collins, Johan Montagnat, Alex P. Zijdenbos, Alan C. Evans, and Douglas L. Arnold

Montreal Neurological Institute
McGill University, Montreal, Canada
{louis,jmontagn,alex,alan,doug}@bic.mni.mcgill.ca
http://www.bic.mni.mcgill.ca

Abstract. Neurodegenerative diseases are often associated with loss of brain tissue volume. Our objective was to develop and evaluate a fully automated method to estimate cerebral volume from magnetic resonance images (MRI) of patients with multiple sclerosis (MS). In this study, MRI data from 17 normal subjects and 68 untreated MS patients was used to test the method. Each MRI volume was corrected for image intensity non-uniformity, intensity normalized, brain masked and tissue classified. The classification results were used to compute a normalized metric of cerebral volume based on the Brain to IntraCranial Capacity Ratio (BICCR).
This paper shows that the computation of BICCR using automated techniques provides a highly reproducible measurement of relative brain tissue volume that eliminates the need for precise repositioning. Initial results indicate that the measure is both robust and precise enough to monitor MS patients over time to estimate brain atrophy. In addition, brain atrophy may yield a more sensitive endpoint for treatment trials in MS and possibly for other neuro-degenerative diseases such as Huntington's or Alzheimer's disease.

1 Introduction and previous work

A number of neuro-degenerative diseases are characterized by brain tissue loss. For example, multiple sclerosis (MS) is a neurological disorder that predominately affects young adults and is associated with recurrent attacks of focal inflammatory demyelination (plaques) that cause neurological impairment, separated by periods of relative stability. It is difficult to evaluate the effect of therapy in clinical trials of MS since it is a complex disease with a high degree of variability in clinical signs and symptoms that vary over time and between individuals. The clinically accepted gold standard measure for *burden of disease* in MS is the Kurztke Expanded Disability Status Scale (EDSS) [1]. Unfortunately, this metric is highly variable between neurologists (large inter-rater variability), is dependent on the timing of the test with respect to the latest exacerbation of the disease and has a variable sensitivity to change depending on the degree

of clinical disability. Taken together, these factors make it difficult to precisely and accurately quantify the overall burden of disease. Therefore, large numbers of subjects (hundreds) are required to participate in clinical trials for new drug evaluation in order to have enough statistical power to detect oftentimes subtle differences between treatment arms.

Our goal has been to develop an objective, automatic, robust image-based method to quantify disease burden in MS. Our interest has turned to central nervous system (CNS) atrophy since histopathological work has confirmed that substantial axonal loss occurs in MS plaques [2] and recent quantitative work has confirmed that CNS atrophy is greater in MS patients than in age-matched normals [3–5].

We propose a fully automated, head-size normalized, brain-volume estimation procedure. The BICCR metric is defined as ratio of brain tissue volume to the total volume enclosed within the intra-cranial cavity. The volumes are derived from the result of a tissue classification process. This metric is similar to the brain parenchymal fraction (BPF) of Fisher [6] where BPF is defined as the ratio of brain tissue volume to total volume enclosed by the brain surface. The main difference is that all extra-cerebral CSF (i.e., CSF between the cortex and dura, in addition to that in the sulci) is included in the BICCR measure. We will show that the BICCR measure is better correlated with disability, and thus may be a better surrogate for disease burden.

2 Methods

2.1 Data

Controls: Seventeen normal healthy controls (age range of 25-61 years) were recruited from the staff, students and research fellows of the Montreal Neurological Institute and McGill community.
Patients: Seventy patients with MS were selected from the population followed in the Montreal Neurological Hospital MS clinic. Forty-eight patients were classified as relapsing-remitting (RR), characterized by recurrent relapses with complete or partial remission (disease duration 0.5 to 24 years, EDSS range 0-5.0, age range 26-58). Twenty-two patients were classified as secondary progressive (SP), characterized by progression in the absence of discrete relapses after earlier RR disease (disease duration 4 to 36 years, EDSS range 3.5-9.0, age range 27-59 years).
MRI acquisistion: All MR data was acquired on a Philips Gyroscan operating at 1.5 T (Philips Medical Systems, Best, The Netherlands) using a standard head coil and a transverse dual-echo, turbo spin-echo sequence, 256x256 matrix, 1 signal average, 250mm field of view, (TR/TE1/TE2 = 2075/32/90 ms) yielding proton density-weighted (PDW) and T2-weighted (T2W) images. Fifty contiguous 3mm slices were acquired approximately parallel to the line connecting the anterior and posterior commissures (AC-PC line).

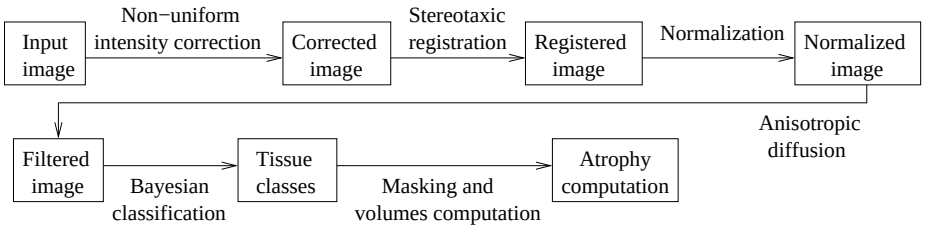


Fig. 1. Diagram of the atrophy computation method stages.

2.2 Data analysis

Atrophy estimation The fully automated method uses MR images to quantify brain atrophy and is based on estimation of the brain to intracranial capacity ratio (BICCR). The method estimates the intracranial, brain parenchymal and CSF volumes and uses these values in a ratio described below.

The technique is voxel-based. Each image voxel is classified as a brain tissue, CSF or background. The number of voxels in each class multiplied by the elementary voxel volume gives an estimate of actual tissue and CSF volumes. As a voxel-based approach, this method requires preliminary processing stages that aim at correcting the image intensities by minimizing the bias and the noise due to the acquisition device. Images are also registered in a common brain-based coordinate space (Talairach) by a linear registration procedure. This ensures that the scale differences between individuals are compensated for and that the resulting atrophy measure is invariant to brain size.

Figure 1 diagrams the atrophy measure stages. The processing stages involve:

Intensity non-uniformity correction. The inhomogeneity of the MR acquisition device magnetic field introduces a bias perceptible in images as a continuous variation of gray-level intensities. The non-uniform intensity correction algorithm [7] iteratively proceeds by computing the image histogram and estimating a smooth intensity mapping function that tends to sharpen peaks in the histogram. The intensities for each tissue type thus have a tighter distribution and are relatively flat over the image volume. Application of this procedure improves the accuracy of the tissue classification stage described below [7].

Stereotaxic registration. Each image is linearly registered in a common Talairach space in order to compensate for size variations between individuals. Moreover, the Talairach-like brain-based coordinate system of stereotaxic space facilitates anatomically driven data manipulation in all processing steps. The target image for stereotaxic registration is a template image built from an earlier study [8] involving the averaging of more than 300 MR images. The registration algorithm proceeds with a coarse-to-fine approach by registering subsampled and blurred MRI volumes with the stereotaxic target [9]. The final data used for subsequent processing is only resampled once to minimize resampling/interpolation artefacts.

Intensity normalization. In preparation for intensity-based classification, each image is intensity normalized to an average PDW (PD-weighted) or T2W (T2-weighted) target volume already in stereotaxic space. An affine intensity mapping is estimated that best maps the histogram of each image onto the template. After normalization, the histogram peaks corresponding to each tissue class have the same value in all images. In conjuction with intensity non-uniformity correction, this step permits data from all subjects to be classified using a single trained classifier (i.e., the classifier does not have to be retrained for each subject).

Cropping. Since the entire cerebrum was not covered by the MRI acquisition in all subjects, the inferior ($z < -22mm$, in Talairach coordinates) and superior ($z > 58mm$) slices were cropped away from both PDW and T2W volumes, cutting off the very top of the brain (above the centrum semi-ovale) and the bottom of the brain (just above the pons). This yielded an anatomically equivalent 80mm thick volume across all subjects that contains most of the cerebrum.

Anisotropic diffusion It has been shown that the application of an edge-preserving noise filter can improve the accuracy and reliability of quantitative measurements obtained from MRI [10, 11]. We have selected anisotropic diffusion, a filter commonly used for the reduction of noise in MRI. This type of filter was pioneered by Perona and Malik [12] and generalized for multidimensional and multispectral MRI processing by Gerig *et al.* [13]. This stage reduces voxel misclassification due to noise and minimizes the speckled appearance sometimes apparent in the resulting classified images.

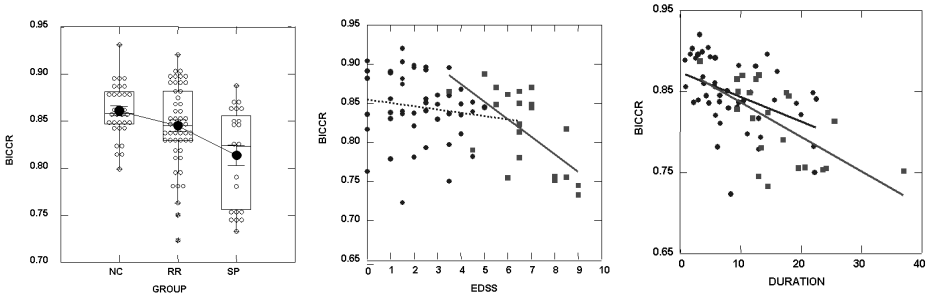
Bayesian classification. A Bayesian classifier [14] is then used to identify all grey-matter (GM), white-matter (WM), cerebrospinal fluid (CSF), lesion (L) and background (BKG) voxels. Prior to classification, the Bayes classifier is trained manually by selecting a set of 20 volumes randomly among all volumes to be processed. From each sample volume, 50 voxels belonging to each class are selected by hand. The resulting 5000 samples (20 volumes $\times$ 50 samples $\times$ 5 classes) were used to compute each class mean intensity and the covariance matrices used in the Bayesian classifier.

Brain masking. Mathematical morphology [15] was used to eliminate the scalp and meninges from further processing. A brain mask was created by applying an *opening* operator (i.e., erosion followed by dilation) to the PDW volume after thresholding at 40% of the mean PDW intensity value. Voxels remaining in the regions of the eyes and nasal sinus were removed using a standard mask in stereotaxic space. The resulting patient-specific brain mask was applied to both the PDW and T2W volumes leaving all voxels within the intracranial cavity.

BICCR computation. After processing, the total volume of voxels in each class was used to define the BICCR metric:

$$\mathrm{BICCR} = \frac{\mathrm{GM} + \mathrm{WM} + \mathrm{L}}{\mathrm{GM} + \mathrm{WM} + \mathrm{L} + \mathrm{CSF}}. \tag{1}$$

It is is important to note that the value of CSF contains all extra-cerebral cerebrospinal fluid within the cropped volume in addition to the ventricular and sulcal components. Similar to the brain parenchymal fraction (BPF) of Fisher



(a) (b) (c)

Fig. 2. Results: (a) box & whisker plot for comparison of BICCR mean values (heavy circles) for NC, RR and SP groups; correlation of BICCR with age (b) and disease duration (c) (RR=black circles, SP=grey squares).

and Rudick [6, 4], the BICCR metric is a ratio and not only represents a size-normalized index of brain atrophy but it also accounts for possible differences in voxel size between scans due to scanner drift.

To determine the reproducibility of the method, 4 healthy volunteers were scanned on 2 separate occasions over a mean period of 222 days. BICCR was computed for each image set. Reproducibility was estimated by computing the coefficient of variation of the repeated measures.

3 Results

The BICCR value for the normal control (NC, n=17) subjects was 86.1 ± 2.8 (mean $\pm$ s.d.). The mean coefficient of variation estimated on scan-rescan tests of 4 normal controls was 0.21%.

Comparison of the mean BICCR values for the NC, RR and SP groups is presented in Figure 2-a. An ANOVA showed a significant difference between groups ($F = 8.885, p < 0.001$). A post-hoc test (Tukey's HSD) showed that BICCR was significantly lower in the secondary progressive group (81.3 ± 5.1) than either the NC group ($p < 0.001$) or the relapsing-remitting group ($84.5 \pm 4.3; p = 0.01$). The Z-score (number of standard deviations from the mean of healthy controls) was -0.673 for RR (not significantly different from NC) and -1.864 ($p < 0.001$) for SP groups. The average absolute percentage of brain tissue lost (compared to normal controls) was 1.8% for RR and 5.6% for SP groups.

We looked at the relationship between BICCR with respect to age, disease duration and EDSS. ANOVA showed no significant differences in age between the NC, RR and SP groups ($F = 1.134, p = 0.327$). As expected, the mean duration of disease of the SP group was significantly greater than that for the RR group (Student's $t = 3.88, p < 0.001$). Also expected, disability (measured by EDSS) was greater for the SP group when compared to that of the RR group ($t = 11.43, p < 0.001$).

For the RR group, BICCR was correlated with disease duration (Spearman $r = -0.523, p < 0.001$), but not with age, nor disability as measured by EDSS (see Figs 2-b and -c). For the SP group, BICCR was correlated with disease duration (Spearman $r = -0.661, p < 0.001$) and EDSS (Spearman $r = -0.649, p < 0.001$) but not with age. When evaluated over all patients with MS (RR and SP combined), BICCR was correlated with EDSS ($r = -0.409, p < 0.01$) and duration ($r = -0.593, p < 0.0001$).

The main difference between the BICCR and BPF metrics is the inclusion of extra-cerebral CSF in the denominator. In a simple test to compare the correlation of disability (measured by EDSS) with BICCR and a measure similar to BPF, we used morphological operators to remove the extra-cerebral CSF voxels from the BICCR metric. When evaluated on 20 SP MS patients, the magnitude of the Spearman's correlation coefficient dropped from -0.638 (BICCR) to -0.574 (modified BICCR).

4 Discussion

We have presented a robust procedure to estimate brain atrophy using a fully automatic technique and have applied it to MRI data from normal controls and patients with MS. We have confirmed that the brains of patients with MS have greater atrophy when compared to normal controls, and that atrophy progresses with the severity and duration of the disease.

Our procedure compares well to the BPF measure of Fisher [6]. The mean BPF and BICCR values are similar for normal controls. However, the BPF method is reported to have a very small intersubject variance when estimated on normal controls (approximately 0.7%). This value is much smaller than the variance for normal controls reported here. This may be due to subject selection and the greater age range for our normal controls.

Another difference between the two techiques is that the classification procedure used in the BPF computation accounts for partial volumes effects between tissue classes, while the BICCR method uses a discrete classification result. While this method should yield an unbiased result for objects that are larger than the voxel size, the BICCR method may underestimate CSF volume in regions that have dimensions on the order of the voxel size, in sulci for example.

The high precision of the BICCR method permits detection of small changes (0.5%) in brain volume (i.e., atrophy) in single subjects over a short period of time (< 1 year). Comparison of BICCR with a BPF-like measure shows that BICCR correlates better with disability, making it possibly a more sensitive surrogate for disease burden. These results have important implications for the design of clinical trials if atrophy is deemed an acceptable surrogate for burden of disease in MS.

The fact that cerebral atrophy is generally correlated with irreversible neurological dysfunction make atrophy an important surrogate to evaluate in MS using state of the art image analysis techniques. Characterization of brain atrophy will yield information complementary to other MR-based measures of focal and diffuse abnormality with varying specificity for underlying pathological changes. Brain atrophy may yield a more sensitive endpoint for treatment trials in MS

and possibly also for other neurdegenerative diseases such as Huntington's or Alzheimer's disease.

Acknowledgements: Funding for this work was provided by the Medical Research Council of Canada.

References

1. J. F. Kurtzke, "Rating neurologic impairment in multiple sclerosis: An expanded disability status scale," *Neurology*, vol. 33, pp. 1444–1452, 1983.
2. B. D. Trapp, J. Peterson, R. M. Ransohoff, R. Rudick, S. Mork, and B. Lars, "Axonal transection in the lesions of multiple sclerosis," *New England Journal of Medicine*, vol. 338, pp. 278–85, 1998.
3. J. Simon, L. Jacobs, M. Campion, et al. "A longitudinal study of brain atrophy in relapsing multiple sclerosis. the multiple sclerosis collaborative research group (MSCRG).," *Neurology*, vol. 53, no. 1, pp. 139–48, 1999.
4. R. Rudick, E. Fisher, J.-C. Lee, J. Simon, D. Miller, and L. Jacobs, "The effect of avonex (ifnβ-1a) on cerebral atrophy in relapsing multiple sclerosis," *Neurology*, vol. 52, pp. A289–290, Apr 1999.
5. M. Filippi, G. Mastronardo, M. A. Rocca, C. Pereira, and G. Comi, "Quantitative volumetric analysis of brain magnetic resonance imaging from patients with multiple sclerosis," *J Neurol Sci*, vol. 158, pp. 148–53, Jun 30 1998.
6. E. Fisher, R. Rudick, J. Tkach, J.-C. Lee, T. Masaryk, J. Simon, J. Cornhill, and J. Cohen, "Automated calculation of whole brain atrophy from magenetic resonance images for monitoring multiple sclerosis," *Neurology*, vol. 52:A352, 1999.
7. J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A non-parametric method for automatic correction of intensity non-uniformity in MRI data," *IEEE Transactions on Medical Imaging*, vol. 17, Feb. 1998.
8. A. C. Evans, D. L. Collins, and B. Milner, "An MRI-based stereotactic atlas from 250 young normal subjects," *Soc.Neurosci.Abstr.*, vol. 18, p. 408, 1992.
9. D. L. Collins, P. Neelin, T. M. Peters, and A. C. Evans, "Automatic 3D intersubject registration of MR volumetric data in standardized talairach space," *Journal of Computer Assisted Tomography*, vol. 18, pp. 192–205, March/April 1994.
10. J. R. Mitchell, S. J. Karlik, D. H. Lee, M. Eliasziw, G. P. Rice, and A. Fenster, "Quantification of multiple sclerosis lesion volumes in 1.5 and 0.5T anisotropically filtered and unfiltered MR exams," *Medical Physics*, vol. 23:115–126; 1996.
11. A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer, "Morphometric analysis of white matter lesions in MR images: Method and validation," *IEEE Transactions on Medical Imaging*, vol. 13, pp. 716–724, Dec. 1994.
12. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 629–639, July 1990.
13. G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data," *IEEE Transactions on Medical Imaging*, vol. 11:221–232; 1992.
14. R. Duda and P. Hart, *Pattern Recognition Scene Analysis.* New York: Wiley, 1973.
15. J. Serra, *Image Analysis, Mathematical Morphology.* London: Academic Press, 1982.