Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Medical Image Analysis 8 (2004) 69–79

MEDICAL IMAGE ANALYSIS

www.elsevier.com/locate/media

The application of serial MRI analysis techniques to the study of cerebral atrophy in late-onset dementia ☆

S.M. Paling [a], E.D. Williams [b], R. Barber [a], E.J. Burton [a], W.R. Crum [c], N.C. Fox [c], J.T. O'Brien [a,*]

[a] *Institute for Ageing and Health, Newcastle General Hospital, Newcastle upon Tyne, UK*
[b] *Regional Medical Physics Department, Sunderland Royal Hospital, Sunderland, UK*
[c] *Dementia Research Group, Institute of Neurology, University College, London, UK*

Received 15 April 2002; received in revised form 10 February 2003; accepted 17 July 2003

Abstract

We have used a serial MR image analysis technique previously developed for studies of cerebral atrophy in early-onset dementia and applied it to a study of late-onset dementia patients with images acquired using a different scanner and scan sequence. Validation and optimisation tests showed that with only small changes to key analysis parameters the technique can successfully be applied to previously untested data with dissimilar image characteristics. The overall accuracy in estimation of cerebral atrophy using the technique was determined to be between 2 and 4 ml (1σ) depending on the conditions during image acquisition. By comparing the results of alternative registration techniques we demonstrate the potential of using of fully automated 9 DOF image registration as an effective and efficient means of correcting for scanner pixel size variations, even in the presence of significant cerebral atrophy. Applied to the late-onset dementia study, patients were found to have significantly increased mean atrophy rates ($p < 0.001$) compared to controls. In general the analysis technique is shown to be a robust, accurate and transferable tool of potential value for future studies of dementia and related neuro-degenerative disorders.

Keywords: Magnetic resonance imaging; Image registration; Cerebral atrophy; Dementia

1. Introduction

The results of cross-sectional MRI studies have shown that an increased level of both global and regional cerebral atrophy is associated with Alzheimer's disease and other forms of dementia (Jack et al., 1998; Barber et al., 2000). Improved methods of investigating cerebral atrophy have subsequently become of considerable interest in the drive to improve our understanding of dementia. They are also of interest in attempts to develop improved methods for differential diagnosis of dementia and for monitoring of disease progression.

Interpretation of cross-sectional studies is commonly limited by inter-patient variability of brain structure and size. Serial studies overcome this variability by comparing a 'baseline' with a 'repeat' scan and observing the changes in a subject's brain over a period of time. A technique of segmentation, registration and comparison of serial 3D MR images for analysis of cerebral atrophy has recently been developed by Freeborough and Fox (Freeborough et al., 1996, 1997; Freeborough and Fox, 1997). This technique was initially developed and validated using locally acquired MR images and successfully applied to investigations of atrophy in a range of neuro-degenerative diseases, in particular early-onset dementia (Fox et al., 1996, 1998, 1999a,b).

This paper describes the results of a detailed validation and optimisation of the Freeborough and Fox technique to ensure its efficacy for application to a new study of cerebral atrophy in late-onset dementia. In being adopted for this study this is the first time the technique has been applied (a) outside the original centre of development, (b) to a late-onset dementia

☆ Electronic annexes associated with this article can be found at doi: 10.1016/j.media.2003.07.04.

* Corresponding author. Tel.: +44-191-256-3323; fax: +44-191-219-5051.
E-mail address: j.t.o'brien@ncl.ac.uk (J.T. O'Brien).

1361-8415/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.media.2003.07.004

cohort, and (c) to data acquired from the particular scanner and scan sequence used in the study (different from that used by the developers).

All serial image comparison techniques are sensitive at some level to variations in image characteristics and quality. Because of this and the need for a high degree of accuracy when studying the subtle structural changes associated with dementia, the detailed evaluation of technique performance described here has been essential for ensuring the validity and accuracy of results in the study. In addition to its relevance here however, the results also provide an important first demonstration of how effectively the Freeborough and Fox technique can be applied to data from a previously untested source with previously untested image characteristics. This will be of particular interest to those considering employing this technique in future studies in different centres, or to those preparing to acquire data for a study in which this or similar techniques will be employed. Elsewhere studies have been carried out comparing the accuracy of the Freeborough and Fox technique with alternative contemporary techniques (Calmon and Roberts, 2000).

We begin by summarising the Freeborough and Fox analysis method and giving details of the data available for our late-onset dementia study. We then describe the validation and optimisation tests carried out and summarize the results of the application of the technique to the late-onset dementia study data. As a part of our discussion we present the results of preliminary investigations into the effects of different aspects of data acquisition on the performance of the analysis technique. In light of our results we consider the transferability and robustness of the technique for future use in different centres and studies.

An 'Electronic Annex' is also provided giving a more visual demonstration of the technique and its application in this study. Analysis of the clinical implications of the outcomes of the late-onset dementia study is presented elsewhere (O'Brien et al., 2001).

2. The Freeborough and Fox analysis technique

The Freeborough and Fox analysis technique is a three stage process designed to be applied to a baseline and repeat 3D T1-weighted MRI scan pair to determine the cerebral atrophy that has occurred during the interval between scans. The three stages of analysis are summarized below. For a more complete description refer to references: (Freeborough et al., 1996, 1997; Freeborough and Fox, 1997; Fox et al., 1996).

2.1. Segmentation

For both the baseline and repeat scans a semi-automated intensity-threshold based process is used to delineate the 3D regions that correspond to either 'brain' or 'non-brain' in each image. Segmentation is carried out using the MIDAS program (Medical Information Display and Analysis System – Freeborough et al., 1997).

2.2. Registration

Three-dimensional rotations, translations and scaling are applied to map the repeat scan onto the baseline scan to eliminate variations in the patient position and image size. Registration is performed using AIR (Automated Image Registration) routines (Woods et al., 1992) which attempt to minimise the difference in signal intensities of corresponding image voxels. Only voxels delineated as 'brain' are used in the calculation so as to minimize the effect of differences in external features on the accuracy of brain tissue alignment.

Registration is carried out in three phases: (i) a coarse determination of the necessary rotation and translation transforms and the position of the approximate global minimum, (ii) a more rigorous iterative adjustment of rotation, translation and global rescaling transformations, and (iii) the application of the resultant transformations to the repeat image. Fast sinc voxel resampling is used throughout phases (ii) and (iii). The final transformations are applied to the entire repeat image, not just to voxels within the segmented brain region.

2.3. Quantification

Assessment of the amount of atrophy that has occurred is determined using the 'Brain Boundary Shift Integral' (BBSI) technique. The registered baseline and repeat scans are 'normalised' to correct for differences in mean brain signal intensity. Then, for all points around the segmented brain boundaries, the technique calculates the volume of brain tissue lost by analysis of the displacements (shifts) of the brain–CSF boundary.

Fig. 1 shows a one-dimensional representation of the baseline and repeat scan intensity profiles and the quantities required to calculate the brain boundary shift (Δx). Essentially the technique determines the size of the shaded area bounded by the baseline and registered repeat intensity profiles (i_{base} and i_{repeat}) and two user specified intensity limits (I_1 and I_2). This area is then divided by the intensity limit interval ($I_1 - I_2$) to give Δx. Mathematically the calculation of area takes the form of an integral and the one dimensional boundary shift is given by

$$\Delta x = \frac{1}{(I_1 - I_2)} \int \big(\text{clip}(i_{\text{base}}(x), I_1, I_2) - \text{clip}\big(i_{\text{repeat}}(x), I_1, I_2\big)\big)\,\mathrm{d}x,$$

where: $\text{clip}(I_{(x)}, I_1, I_2) = I_{(x)}$ for $I_1 > I_{(x)} > I_2$, I_1 for $I_{(x)} > I_1$, and I_2 for $I_{(x)} < I_2$.

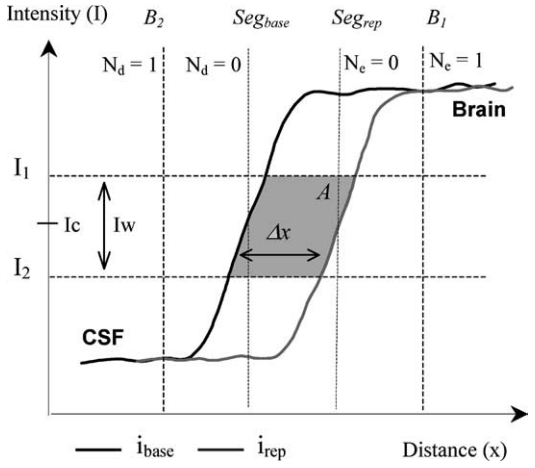

Fig. 1. A one-dimensional representation of baseline and repeat scan intensity profiles and the parameters required to determine cerebral atrophy using the Brain Boundary Shift Integral (BBSI). At all points around the brain boundary the BBSI technique measures small shifts in the brain–CSF boundary (Δx) by analysis of the shaded area (A) enclosed by the baseline and repeat scan intensity profiles (i_{base} and i_{rep}) and the intensity window limits (I_1 and I_2). B_1 and B_2 are the boundary limits defining the region over which the BBSI calculation is carried out. Seg_{base} and Seg_{repeat} are the original baseline and repeat brain–CSF boundaries determined during segmentation. (I_c, I_w, N_d and N_e are user defined 'calibration' parameters used to set the intensity window limits I_1 and I_2, and the brain boundary limits B_1 and B_2.)

Extending the above technique to three dimensions enables the change in brain volume to be determined from observed shifts in the brain boundary in the x, y and z directions. In practice, the integration is calculated numerically by summation of the difference in intensity values of all voxel elements within the boundary region (again using the I_1 and I_2 intensity limit cuts and normalisation). The resulting volume of brain tissue lost can be expressed either as a volume (e.g. in ml) or as a percentage reduction of total brain volume.

An electronic annex submitted with this paper provides a visual overview of the Freeborough and Fox technique including example MR image pairs at key stages of the analysis procedure. Also provided in the annex is a more detailed schematic of the BBSI calculation procedure and a previously unpublished visual demonstration of the BBSI algorithm in action.

3. Subjects and scans

The study of cerebral atrophy for which the above analysis technique has been employed forms an important part of a wider series of investigations of late-onset dementia by the Institute for Ageing and Health at Newcastle General Hospital. For the MR imaging component of the study, two T1-weighted 3D MRI scans were acquired at approximately 1 year intervals for 10 patients with probable Dementia with Lewy Bodies (DLB), eight with Alzheimer's Disease (AD), 10 with Vascular Dementia (VaD) and 20 age-matched controls (Con). Patients were recruited from referrals to the local old age psychiatry service and were diagnosed using currently accepted consensus criteria (McKhann et al., 1984; Roman et al., 1993; McKeith et al., 1996). The study was approved by the Newcastle and North Tyneside Joint Ethics Committee and written informed consent was obtained from all participating subjects. All images were acquired at the Royal Victoria Infirmary at Newcastle, UK using a Siemens 1.0T Impact Expert System scanner using a spoiled gradient echo sequence (FSPGR) with contiguous $1 \times 1 \times 1$ mm voxels, T_R: 11.4 mS, T_E: 4.4 mS, T_i: 400 mS, flip: 15°. Images were subsequently transferred to a Sun Ultra 30 workstation and processed using the analysis techniques described above.

To enable a thorough validation and optimisation of the analysis technique applied to the study a series of same-day repeat scans were also acquired for two healthy volunteers; one elderly (aged 71) and one younger (aged 25). On a particular day, six repeat scans were obtained for each volunteer, using the same scanner and scan sequence as used for the late-onset dementia study. Each of these scans was compared with the other five, generating 15 scan-pair combinations for later analysis.

4. Technique validation and optimisation

Validation and optimisation were considered essential before the Freeborough and Fox analysis technique could be applied to data with different origins and image characteristics from those for which it had been originally developed. Earlier studies using the technique had been carried out on predominantly younger subjects than those used in the late-onset dementia study (e.g., a mean age of <55 in a study of early-onset Alzheimer's disease (Fox et al., 1996)). Also, images previously used had been acquired using a different and higher field strength scanner (a 1.5 T GE Signa), slightly different scan sequences (although T1 weighted FSPGR in all cases) and a different voxel size ($1.0 \times 1.0 \times 1.5$ mm). Each of these differences could result in significant changes in the performance of the analysis technique when applied to the study of late-onset dementia.

Four aspects of the analysis technique were considered important for validation and optimisation in preparation for the clinical study. These were BBSI calibration, BBSI gain and linearity, image scaling and overall measurement accuracy. Each of these are described in detail below. The first two (BBSI calibration, gain and linearity) were similar to investigations that had already been carried out during the original development of the analysis technique (Freeborough et al., 1997). The image scaling investigations carried out are new to this study and demonstrate the potential of an

alternative method for correcting for MR scanner image scaling variations to that previously used in the technique. Finally, the overall accuracy of the Freeborough and Fox technique was assessed by methods both similar to and different from those previously employed and reported. These results are of particular interest as they demonstrate how the overall performance of the technique varies when applied to data from a previously untested source.

4.1. BBSI calibration

The analysis technique requires that four 'calibration parameters' (I_c, I_w, N_d and N_e) are set to ensure the BBSI is calculated appropriately for the images being analysed. I_c and I_w (expressed as fractions of the mean brain intensity) represent the centre and width of the intensity window that defines the range of voxel intensities over which the CSF-brain boundary shift is assessed (the window bounded by I_1 and I_2 in Fig. 1). N_d and N_e (expressed as number of voxel widths) are the number of dilations and erosions applied to the boundaries of the union (i.e., the outermost boundary) and intersection (i.e., the outermost boundary) of the segmented brain boundaries (Seg_{base} and Seg_{rep}). These define the region around the brain boundary over which the BBSI analysis is applied (the region bounded by B_1 and B_2). Earlier studies found appropriate values of these parameters to be: $I_w = I_c = 0.5 \times$ the mean brain intensity, and $N_e = N_d = 1$ voxel widths. However, for application to data of different image characteristics in the late-onset dementia study a new search for optimum values was required.

As for the earlier studies, optimum values for the BBSI calibration parameters were determined by inspection of 'calibration curves' in which the individual user selected parameters were varied and plotted against the resulting BBSI output. Calibration curves were generated for each subject in the late-onset dementia study. In addition to enabling an assessment of appropriate and optimum calibration parameters the curves provide a valuable means of checking how the BBSI analysis was performing throughout the study. ('Appropriate' and 'optimum' calibration values in this case mean those that both ensure the BBSI summation is correctly calculated, and that provide the best overall atrophy measurement accuracy when applied to all patients in the study).

A typical set of calibration curves obtained for one dementia patient used in the late-onset dementia study is shown in Fig. 2. Figs. 2(a) and (b) show the effects of changing first N_e and then N_d on the BBSI when I_w is very small and I_c is varied from 0.3 to $0.9 \times$ the mean brain intensity. The curves show that for this scan pair BBSI reached a maximum when I_c was about 0.65. The curves also show a larger decrease in BBSI for values of

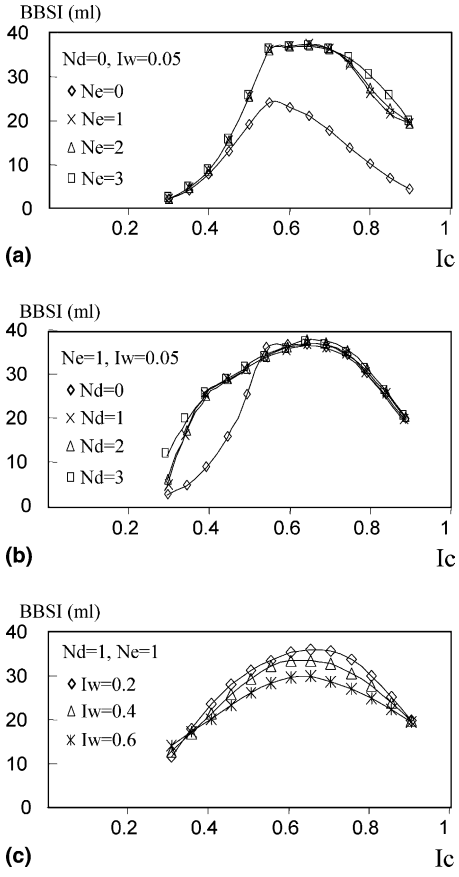

Fig. 2. Example of the BBSI 'calibration curves' generated during analysis of the baseline and repeat scan pair of a patient in the late onset dementia study. The curves show the effect of varying the parameters: (a) N_e, (b) N_d and (c) I_w on the BBSI estimation of volume of brain tissue lost due to atrophy. I_c and I_w are the centre and width of the BBSI intensity window (bounded by the intensity levels I_1 and I_2 defined previously). All intensity values are quoted as fractions of mean segmented brain intensity (see text for further explanation of terms).

I_c deviating from 0.65 when N_d and N_e were less than one voxel width, indicating that at least one erosion and one dilation needed to be applied to the segmented boundary regions to be sure the brain–CSF boundaries of both images were adequately covered during BBSI estimation.

Fig. 2(c) shows the variation in BBSI output when N_e and N_d were kept constant and both I_w and I_c were varied. As noted in earlier studies BBSI output was seen to be consistently lower for larger values of I_w. This effect is caused by a greater proportion of 'imperfect' low intensity difference boundaries being included in the BBSI calculation when the intensity window width is large (e.g., boundaries where grey matter is not adjacent to CSF or where the grey matter intensity is relatively low). Calibration curves were also seen to be consistently broader for larger values of I_w. A broad calibration curve is potentially advantageous for reducing uncertainties due to inter-subject variations if common

I_c and I_w values are to be used for all of the subjects in the study. If such an approach is taken then the mean of the individual I_c values found to give the maximum BBSI in each subject is a reasonable common I_c value to use. Choice of the optimum I_w involves a more arbitrary trade-off between BBSI gain and robustness to inter-subject variations.

Some anomalous calibration curves were also noted. Occasionally peaks or troughs were seen at the extremities of the N_e and N_d calibration curves which corresponded to the approximate intensity levels of grey matter and CSF. On closer visual inspection these characteristics were found to be most likely caused by differences in the level of tissue blurring (caused by patient movement) in the baseline and repeat images, or by the presence of occasional marked image intensity non-uniformity. Before deciding upon the optimum parameters to use in the study tests were carried out to see if it was better to use a small intensity window (I_w) to stay clear of the regions where the anomalies in calibration curves were most frequent. The scan pairs of all control patients and a set of same-day repeat scans were processed using values of I_w ranging from 0.2 to $0.5 \times$ the mean brain intensity. No significant difference in the spread of results were found, showing that any adverse effects of partially including these anomalous regions were negligible with respect to the results of this study.

After consideration of all of the calibration curves obtained the final BBSI parameter values chosen for the late-onset dementia study were: $N_d = 1$, $N_e = 1$, $I_c = 0.55$ and $I_w = 0.5$. Without the above tests having demonstrated clear incentives for choosing a particular value of I_w, the value of 0.5 was chosen partly for consistency with previous studies.

4.2. BBSI gain and linearity

Whenever the BBSI technique is applied to the analysis of baseline-repeat scan pair the presence of numerous localised imperfections of the brain–CSF boundary in the images causes the technique to return a lower estimate of atrophy than is actually present. Such imperfections (present in all images) include transitions from brain directly to (for example) dura instead of CSF, or true brain–CSF transitions in regions where the difference in the tissue signal intensities is markedly lower than usual. In regions where the boundary intensity differences are negligible no boundary shifts are detected and any contributions to the overall atrophy are missed. In regions where boundary intensity differences are significant but small (and in particular where the brain/non-brain levels fall within the BBSI intensity limits I_1 and I_2 described above) the conditions for the appropriate application of the technique are not satisfied and an incorrect and commonly smaller value of atrophy is returned. Summed over all regions of the brain this results a moderate but consistent underestimation of atrophy, the level of underestimation being dependant on the BBSI calibration parameters chosen and the characteristics of the images being studied. To account for this underestimation of atrophy it is important that a BBSI 'gain value' is determined to enable the conversion from the recorded BBSI output to the true atrophy level.

Another important BBSI technique performance characteristic is measurement 'linearity'. Either the BBSI gain should be constant for all levels of atrophy (i.e. linear) or variations in BBSI gain should be understood and corrected to ensure accurate and unbiased measurement for all scan pairs studied.

As in earlier studies the BBSI gain and linearity were assessed by observing the response of the analysis technique to 'simulated atrophy', introduced by applying different levels of three-dimensional (3D) scaling to a baseline or a repeat scan. These scans were subsequently processed in the usual way, but without using any voxel scaling correction, so that the simulated atrophy would not be automatically eliminated during registration. A wide range of simulated atrophy values were applied to each scan pair, the results enabling a precise assessment of BBSI gain and linearity for each subject studied. Gain and linearity were also studied by comparison of the BBSI output with differences in the un-scaled segmented brain volumes between scan pairs. The difference in segmented brain volumes is an independent measure of atrophy that is expected to be less accurate because of inherent variability in the segmentation process. Comparison of the two values provides a means of testing the BBSI gain and linearity in response to real atrophy in the subjects to be studied.

Results of BBSI gain and linearity tests are shown in Fig. 3. Fig. 3(a) shows the result of plotting BBSI output against applied simulated atrophy for one subject of the late-onset dementia study. The results are typical of subjects tested and show the BBSI technique to be highly linear and with (in this case) a gain of approximately 0.76 using $I_w = 0.5$. Subsequent assessment of the full late-onset dementia group showed the mean BBSI gain to be 0.73 (± 0.01) with less than 8% (1σ) variation from patient to patient. Fig. 3(b) shows BBSI output plotted against the difference in volume of the segmented brain regions for a number of the subjects in the late-onset dementia study. The results indicate that the BBSI response to real atrophy is also linear. The relatively large scatter seen is principally due to the inherent variability and inaccuracy of brain volumes determined by the semi-automated segmentation process (and for clarity, only results for subjects for which scan pair segmentation was judged to be of high quality are shown). The mean BBSI gain value determined from this data is 0.66 (± 0.10), which is in agreement with the simulated atrophy results described above.

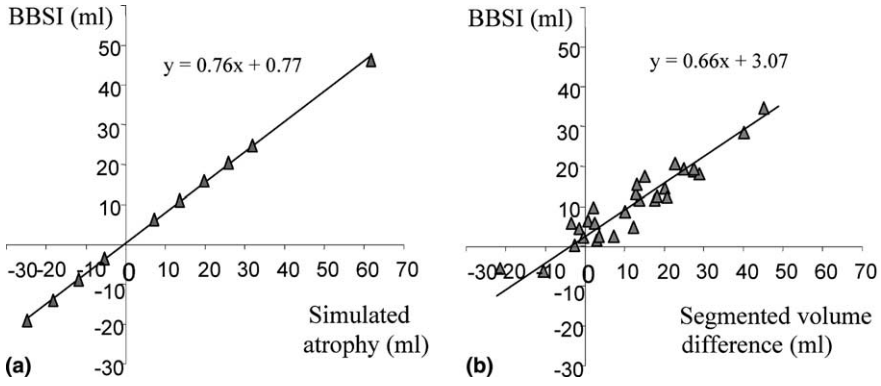

Fig. 3. Investigations of BBSI linearity and gain. (a) The response of BBSI to simulated atrophy for one subject in the late-onset dementia study. Simulated atrophy was introduced by applying 10 different levels of 3D scaling to one of the images of the subject's scan pair. (b) BBSI atrophy measurements plotted against atrophy determined by the difference in segmented brain volumes for subjects in the late-onset dementia study.

Both assessment methods indicate the BBSI technique provides a linear measure of atrophy. The simulated atrophy results however, provide the more precise measure of mean BBSI gain and so the value of 0.73 was subsequently adopted as the gain value for all subjects in the study.

4.3. Image scaling correction

Without correction, any variations in scanner voxel size between the baseline and repeat scan would erroneously be interpreted and brain tissue loss or gain by the Freeborough and Fox analysis technique. An independent quality assurance investigation of the scanner used in this study (Firbank et al., 2000) showed voxel size variations over a period of year to be in the order of 0.5% (1σ), with no more than 1% scaling error in any given axis. While such variations are common (and within manufacturers specifications), without correction they would directly result in an additional atrophy measurement uncertainty of 0.5%, which itself is comparable to the annual levels of atrophy previously observed in controls and some dementia subjects.

Tests carried out during the original development of the Freeborough and Fox technique showed that using intra-cranial volume measurements to determine 3D scaling correction factors during the image registration stage of the technique could significantly reduce the effects of variations in scanner voxel size (Freeborough et al., 1996). Unfortunately, cranial cavities were difficult to outline on the images used in the late-onset dementia study and the resulting voxel scaling correction factors were found to be inaccurate and unreliable. As an alternative we investigated the use of a fully-automated nine degrees-of-freedom (9 DOF) image registration technique in which 3D image scaling correction values are determined during the automated voxel-based registration (in addition to the 3D translations and rotations already determined by the algorithm).

Our primary areas of concern when using the automated 9 DOF registration technique for the study were: (a) would the technique accurately correct for scanner voxel size variations without introducing a greater level of uncertainty into the assessment of atrophy? (b) Would the technique be biased by real atrophy in the images? More specifically: would it mistake real atrophy for scanner voxel size changes and tend to over-correct during re-scaling? A number of tests were carried out to address these concerns.

To test whether 9 DOF registration would appropriately correct for scanner voxel variations and give an overall improvement in technique accuracy, atrophy measurements were determined both with and without automated scaling correction (i.e., using 9 DOF and 6 DOF registration respectively) and results were compared for both the same-day repeat scans and all control subjects in the late-onset dementia study. In both sets of data one would expect little or no true atrophy and the resulting distribution of measured atrophy results provides an indication of the level of measurement uncertainty present. Any reduction in spread observed using 9 DOF instead of 6 DOF technique would also indicate both that scanner voxel size variations were significant for the study and that the automated scaling technique was resulting in a reduced uncertainty in atrophy measurement as required.

For both the same-day repeat scans and the control scans the 9 DOF registration was seen to result in a tighter grouping of atrophy results than the 6 DOF registration. Statistically however, the reduction in spread was not found to be significant. (For the same-day repeat scans the spread in atrophy results (1σ) was 4.1 ml using 9 DOF and 5.5 ml for 6 DOF registration. For the control subject scans the spread was 8.0 ml for 9 DOF and 9.0 ml for 6 DOF registration.)

Further visual inspection of the scans used in these tests indicated that 9 DOF registration was adjusting for real voxel size variations as desired, but the resulting

improvement in measurement uncertainty was not sufficiently large to be seen above other sources of variation in the results. The relatively small reduction in spread observed for the same-day scans may have been because the most significant scanner voxel size variations actually occur over longer time periods than a single day. The lack of a significant reduction in the control subject group (with approximately 1 year scan intervals) may in turn have been due to the domination of the additional sources of variability for the data set, including real variations in the levels of atrophy of the subjects studied.

To investigate any bias of the 9 DOF registration technique in the presence of genuine atrophy tests were carried out to see if there was any correlation between the 3D scaling values applied during registration and the level of atrophy subsequently determined for each scan pair. Any correlation would suggest that the values of 9 DOF scaling applied were dependant not only on scanner voxel size variations but also on the level of atrophy present: without correction the 9 DOF technique could then not be relied upon to give an accurate measurement of atrophy.

Fig. 4 shows the result of plotting the 9 DOF registration scaling values against measured atrophy for all subjects used in the late-onset dementia study. The observed scaling values range from between 0.985 and 1.006 and have a mean and standard deviation of 0.998 and 0.005 respectively. Importantly, no significant correlation between the applied scaling values and measured level of atrophy is seen. This shows that the 9 DOF voxel size correction method was not significantly biased by the presence of atrophy. In addition, the overall spread in scaling values is in good agreement with the voxel size variations determined during independent phantom studies of the scanner performance. This further supports the evidence that 9 DOF registration was responding to real voxel variations as required and not introducing a significant number of spurious scaling changes that might result in an overall increase in atrophy measurement uncertainty.

To further test the performance of the 9 DOF registration technique, the applied scaling correction values were analysed with respect to the calendar dates on which baseline and repeat scans had been acquired. If (as might be expected) scanner variations occurring in a single day were smaller than those occurring over longer time periods, then quite similar applied scaling values would be expected for subjects whose year-apart scans were acquired on the same days. Conversely, a larger spread of values would be expected for those subjects whose year-apart scans were acquired on different days because of the greater expected scanner voxel size variations.

Our attempts to detect such an effect are shown in Fig. 5. The figure shows the results of plotting the applied scaling values for all patients in the study in order of the mean scan date of each baseline and repeat scan pair. The data shown have been separated into two categories: subjects with unique combinations of baseline and repeat scan dates (unmatched subjects) and subjects with scan dates common to at least one other subject (matched subjects). The grouping of a number of matched subject groups appears closer than that of the unmatched subjects as would be expected if 9 DOF registration was accurately adjusting for voxel size variations. For a few of the matched subject groups

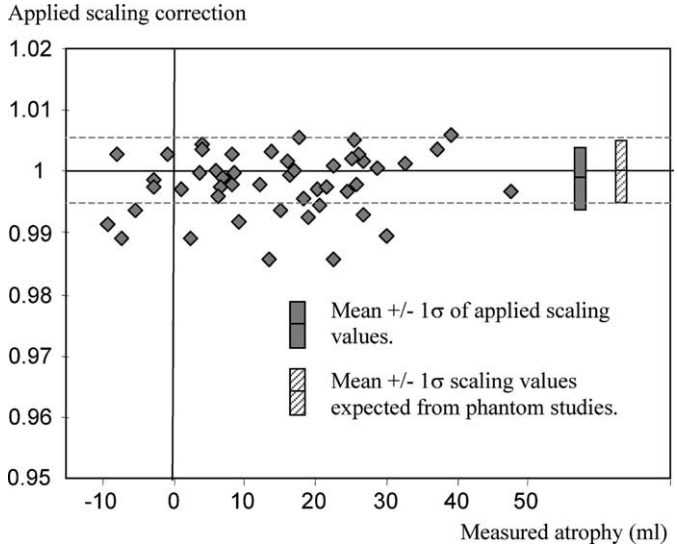

Fig. 4. The performance of the 9 DOF registration voxel size correction technique. Applied scaling correction values are plotted against the measured level of atrophy for all subjects in the late-onset dementia study. Also shown are the mean $\pm 1\sigma$ applied scaling values, and the mean $\pm 1\sigma$ scaling values required to correct for the level of voxel size variations observed during independent phantom studies of the scanner. (A value of e.g. 0.99 in this plot means the 9 DOF registration has reduced the size of the repeat scan by 1%.)

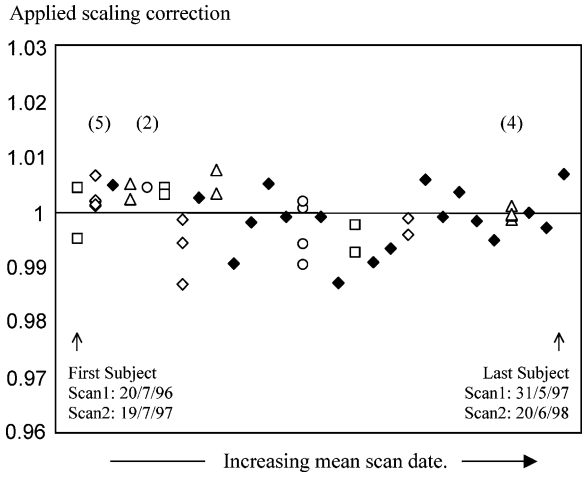

Fig. 5. The variation of applied 9 DOF scaling correction values over time. The applied 9 DOF scaling values of all subjects in the late-onset dementia study are plotted in order of mean scan date. Filled symbols represent subjects with unique baseline and repeat scan dates (unmatched subjects), unfilled symbols represent subjects for which at least one other subject has identical scan dates (matched subjects). Bracketed numbers show the total number of matched subjects where applied scaling values are so similar that individual symbols are obscured.

however the spread is seen to be just as large as that of the unmatched subjects. The result suggests that either same-day voxel size variations are occasionally as large as those seen over longer time periods, or on occasion the 9 DOF scaling value determination process was significantly affected by factors other than variations in scanner voxel size. Close inspection of those matched subject's scans revealed greater than usual levels of image blurring due to patient movement and it appears likely that at least part of the observed spread could be attributed to the effects of poor image quality.

Overall the results of tests described above clearly suggest that for this study, 9 DOF registration adequately and appropriately corrected for scanner voxel size variations without being significantly biased in the presence of true atrophy. Unfortunately, because of relatively large inherent variations in the data used to test the technique it was not possible to demonstrate with statistical significance that 9 DOF registration reduced the overall level of uncertainty in the measurement of atrophy. However, since in no instance was 9 DOF registration seen to introduce more uncertainty than using no correction method at all, automated 9 DOF registration was considered to be the appropriate technique to use for the remainder of the study.

The finding that for 9 DOF registration appears to appropriately correct for scanner voxel size variations without being biased by the presence of true atrophy may at first seem counter intuitive. A priori, one instead might expect there to be a tendency for the technique to cancel out atrophy during rescaling, the reduction of brain size being mistaken at some level for a scanner voxel size reduction. The reason this is found not to be the case is understood to be due to the high level of complexity of the shape of the object being transformed. If the brain were simply a uniform solid sphere then one would indeed expect uniform and diffuse atrophy to be entirely cancelled out during rescaling. If however one imagines a shape of increasing complexity, for example with holes (like ventricles) and complex limbs (like sucli), then diffuse atrophy results in a more complex, nonlinear shape change and the resulting voxel position differences can no longer be corrected for by simple linear 3D rescaling. It follows that as complexity increases the level of 'inappropriate' rescaling in the presence of diffuse atrophy becomes smaller and smaller, to the point where (as seen in this study) the effect becomes insignificant.

4.4. Assessment of accuracy

Finally, tests were carried out to determine the overall accuracy of the optimised technique when applied to the late-onset dementia data. Tests during earlier studies using the technique had estimated the overall accuracy in measurement of whole brain atrophy to be <2 ml (1σ) (Fox and Freeborough, 1997). The comparison of this figure with the accuracy achieved in the late-onset dementia study was of particular interest as were any indications of which aspects of image acquisition and processing would most affect atrophy measurement accuracy.

The atrophy measurement accuracy can be determined in a number of ways, although the fact that different error factors may be more important to different patient scans and the expected presence of significant brain size changes of even the control subjects makes a precise value for the overall accuracy in the study difficult to obtain. The most appropriate method for this study was considered to be by analysis of the spread of atrophy results for the 15 same-day repeat scan pairs of the elderly and younger controls. The spread of true atrophy levels for these subjects should be negligible and all deviations from zero atrophy should therefore represent the level of measurement uncertainty in that particular scan series.

The most likely sources of uncertainty in the technique were considered to be; random uncertainties in the application of BBSI, errors in image registration, image artefacts (e.g., image noise or movement artefacts) and genuine temporal fluctuations in subject brain volume. The scan series which were subject to most of these were those of the elderly control scanned at approximately 1-h intervals throughout one day. These were therefore used to determine the most realistic overall measurement uncertainty for the late-onset dementia study. In contrast, the 'ideal' or minimum uncertainty attainable

using the scanner, scan sequence and analysis technique was determined from repeat scans of the younger control with no significant changes in head position, generally better defined brain–CSF boundaries and with no significant time interval between scans.

Using the BBSI calibration values described above, the 9 DOF registration analysis of the same-day repeat scan pairs of the elderly volunteer resulted in a spread in atrophy values of $(1\sigma) = 4.1$ ml (± 1.5 ml 95% confidence limits). The overall measurement accuracy for the study is therefore considered to be: Δatrophy $(1\sigma) \approx 4$ ml. This is significantly higher than the figure of <2 ml quoted in the previous studies. The analysis of the repeat scans of the younger volunteer showed the minimum uncertainty attainable using this scanner, scan sequence and analysis technique to be: Δatrophy $(1\sigma) \approx 2.0$ ml (± 0.7 ml 95% confidence limits). This is comparable with the accuracy quoted in the previous studies, but with the level of control applied during image acquisition the result is considered a lower limit of what could practically be achieved in serial study of elderly subjects such as this.

5. Application to the late-onset dementia study group

The results of the final analysis of total cerebral atrophy for all of the subjects in the late-onset dementia study are shown in Fig. 6. Atrophy measures are expressed as ml of brain volume lost per year (corrections have been made for variations in individual baseline and repeat scan intervals). Both individual atrophy measures and mean atrophy rates for each dementia group are shown.

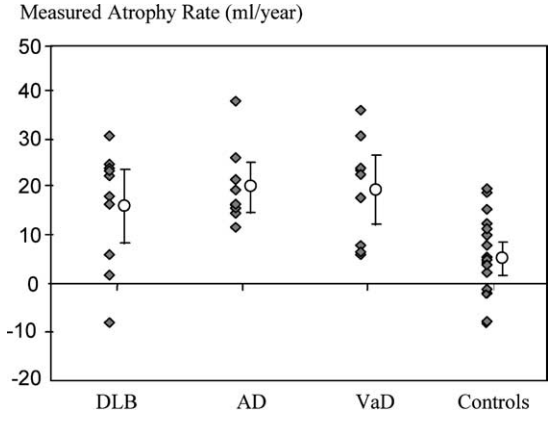

Fig. 6. Comparison of the measured rates of cerebral atrophy (i.e., atrophy occurring over approximately 1 year) for all subjects in the late-onset dementia study. Atrophy rates have been determined using the optimized and validated analysis technique including 9 DOF registration voxel size correction. Filled symbols show individual atrophy results, unfilled symbols and error bars show mean atrophy rates $\pm$95% confidence limits.

The mean $\pm$ 95% confidence limits for control and patient groups are: DLB 15.8 ± 7.6 ml; AD 19.9 ± 5.2 ml; VaD 19.4 ± 7.1 ml; controls 5.2 ± 3.5 ml. The application of ANOVA and post hoc tests to these results show dementia subjects to have significantly increased mean atrophy rates ($p < 0.001$) compared to controls, with no significant difference in the mean atrophy rates between dementia groups. A considerable overlap in the dementia and controls groups is seen. This, and the somewhat surprisingly large spread in the measurement of atrophy for control subjects, is partly due to the estimated overall measurement uncertainty of Δatrophy $(1\sigma) \approx 4$ ml.

6. Discussion

Overall, the results of our studies show that the Freeborough and Fox analysis technique can be successfully applied to data of dissimilar and previously untested image characteristics. Interestingly however, even after optimisation and validation we found the level of accuracy of the technique (applied to our late-onset dementia study data) to be slightly though significantly lower than that specified in the earlier studies.

To attempt to address this a selection of further same-day repeat scan tests have since been carried out to ascertain a) how this difference may have arisen, b) what the principle sources of uncertainty are, and c) what (if anything) could be done to improve the accuracy of the technique in the future. Unfortunately a full and rigorous test of all potential variables has not been possible because of the prohibitive time & expense required to acquire multiple same-day scans under many different conditions. The results obtained however, suggest that the factors that most affect the accuracy of atrophy measurement include patient movement during scanning and inconsistency in patient positioning from one scan to the next. Potentially important factors that appear to have little effect were moderate image non-uniformity, moderate changes in scan sequence (though still with the same sequence used for both baseline and repeat scan) and natural fluctuations in brain size occurring throughout the day.

These above findings can only be considered preliminary and anecdotal and more work is required to more precisely determine the factors most significantly influencing the accuracy of the technique. Although uncertain therefore, we believe it most likely the difference between the accuracy observed for this and previously reported studies is due to there having been more consistent patient positioning or a reduced level of patient movement in the earlier studies. At this stage (and until further studies have been carried out) it is these factors we would recommend particular care be given to when acquiring data for a study in which the Freeborough and Fox analysis technique is to be used.

Another important outcome of our studies has been to demonstrate that automated 9 DOF registration can be used as an efficient means for correcting scanner voxel size variations without being significantly biased by the presence of cerebral atrophy. As discussed earlier, although perhaps initially counterintuitive, the lack of bias of the technique in the presence of atrophy can be explained by the relatively high complexity of the shape of the brain and the inability of a simple linear rescaling transformation to correct for voxel position differences caused by the correspondingly complex pattern of diffuse atrophy.

This encouraging finding raises the questions of whether 9 DOF registration can be confidently used as alternative to traditional methods for adjusting for scanner voxel size variations such as using information from quality control phantoms or measurements of size invariant structures such as inter-cranial volumes. A difficulty is that for any new study using the technique some level of 'inappropriate' compensation for voxel size change due to atrophy may still be expected to occur and exactly how much and how significant this would be is difficult to predict. The level may vary with the size and shape of the objects being studied or with general image characteristics such as voxel size, contrast, scanner type and sequence, etc. Also, as indicated in this study, the extent to which 3D rescaling actually achieves its purpose of adjusting for true scanner variations may be affected by image quality factors such as poor contrast, motion artefacts, image non-uniformity, etc.

Overall, we believe that in some circumstances the 9 DOF registration technique will prove to be a sufficiently accurate and unbiased means of adjusting for scanner voxel size variations even in the presence of significant cerebral atrophy. At this early stage however, we recommend that comprehensive validation tests are carried out when considering employing the technique in future studies.

7. Conclusions

By carrying out detailed performance studies we have shown that with minor adjustments of key analysis parameters the Freeborough and Fox analysis technique can successfully be applied to data from a previously untested source with different image characteristics. The overall accuracy in estimation of cerebral atrophy using the technique was found to be between 2 and 4 ml (1σ) depending on the level of care taken during image acquisition. More work is required to precisely determine the most significant factors influencing the accuracy of the technique, however preliminary investigations suggest the reduction of patient movement and consistency of patient repositioning are worthy of particular attention during image acquisition.

By comparing the results of alternative registration techniques we have shown that the use of fully automated 9 DOF image registration shows potential as an efficient means of correcting for scanner voxel size variations without introducing additional uncertainty or being biased by the presence of cerebral atrophy. We believe that in some circumstances this technique could have significant advantages over traditional voxel size correction methods. At present however, the factors influencing the performance of the technique in this role are not well understood and further tests are recommended when considered for future studies.

Applying the Freeborough and Fox technique to the late-onset dementia study, a group of Dementia with Lewy Bodies, Alzheimer's Disease and Vascular Dementia patients were found to have significantly increased mean atrophy rates ($p < 0.001$) compared to controls, with no significant difference in the mean atrophy rates between dementia groups.

Overall we find the Freeborough and Fox analysis technique to be a generally robust, accurate and transferable tool for determining cerebral atrophy from serial MR images and expect it to be of continued value in the study of dementia and related neuro-degenerative disorders. The detailed technique validation and optimisation tests described were, however, found to be both useful and necessary. Similar tests are therefore recommended for others considering applying the technique to new studies in different centres.

Note

A series of previously unpublished images and movies are supplied (at doi: 10.1016/j.media.2003.07.04) which better illustrate the Freeborough and Fox analysis technique in action. (Detailed explanation of the technique is given in Section 2 of the article text.)

Any comments or questions regarding these images, the work discussed in the article or any other aspects of serial imaging and late onset-dementia research are very welcome. Please contact the authors at S.M.paling@shef.ac.uk (S.M. Paling) or J.T.O'Brien@ncl.ac.uk (J.T. O'Brien).

Acknowledgements

The authors thank the Medical Research Council, UK, for financial support.

References

Barber, R., Ballard, C., McKeith, I.G., Gholkar, A., O'Brien, J.T., 2000. MRI volumetric study of dementia with Lewy bodies: a

comparison with AD and vascular dementia. Neurology 54, 1304–1309.

Calmon, G., Roberts, N., 2000. Automatic measurement of changes in brain volume on consecutive 3D MR images by segmentation propagation. Magn. Reson. Imaging 18, 439–453.

Firbank, M.J., Harrison, R.M., Williams, E.D., Coulthard, A., 2000. Quality assurance for MRI: practical experience. Br. J. Radiol. 73, 376–383.

Fox, N.C., Freeborough, P.A., 1997. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. J. Magn. Reson. Imaging 7, 1069–1075.

Fox, N.C., Freeborough, P.A., Rossor, M.N., 1996. Visualisation and quantification of rates of atrophy in Alzheimer's disease. Lancet 348, 94–97.

Fox, N.C., Warrington, E.K., Seiffer, A.L., Agnew, S.K., Rossor, M.N., 1998. Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. A longitudinal prospective study. Brain 121, 1631–1639.

Fox, N.C., Scahill, R.I., Crum, W.R., Rossor, M.N., 1999a. Correlation between rates of brain atrophy and cognitive decline in AD. Neurology 52, 1687–1689.

Fox, N.C., Warrington, E.K., Rossor, M.N., 1999b. Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease [letter]. Lancet 353, 2125.

Freeborough, P.A., Fox, N.C., 1997. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans. Med. Imaging 16, 623–629.

Freeborough, P.A., Woods, R.P., Fox, N.C., 1996. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. J. Comput. Assist. Tomogr. 20, 1012–1022.

Freeborough, P.A., Fox, N.C., Kitney, R.I., 1997. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. Comput. Methods Programs Biomed. 53, 15–25.

Jack, C.R., Petersen, R.C., Xu, Y., et al., 1998. Rate of medial temporal lobe atrophy in typical ageing and Alzheimer's disease. Neurology 51, 993–999.

McKeith, I.G., Galasko, D., Kosaka, K., et al., 1996. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the consortium on DLB international workshop. Neurology 47, 1113–1124.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–944.

O'Brien, J.T., Paling, S., Barber, R., Williams, D.W., Ballard, C., McKeith, I.G., et al., 2001. Progressive brain atrophy on serial MRI in dementia with Lewy bodies, AD and vascular dementia. Neurology 56, 1386–1388.

Roman, G.C., Tatemichi, T.K., Erkinjuntti, T., et al., 1993. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop [see comments]. Neurology 43, 250–260.

Woods, R.P., Cherry, S.R., Mazziotta, J.C., 1992. Rapid automated algorithm for aligning and reslicing PET images. J. Comput. Assist. Tomogr. 16, 620–633.