NEURO

C. Farrell
F. Chappell
P. A. Armitage
P. Keston
A. MacLullich
S. Shenkin
J. M. Wardlaw

Development and initial testing of normal reference MR images for the brain at ages 65–70 and 75–80 years

C. Farrell · F. Chappell ·
P. A. Armitage · P. Keston ·
J. M. Wardlaw (✉)
Division of Clinical Neurosciences,
Western General Hospital,
Bramwell Dott Building, Crewe Rd,
Edinburgh, EH4 2XU, UK
e-mail: joanna.wardlaw@ed.ac.uk
Tel.: +44-131-5372943
Fax: +44-131-3325150

A. MacLullich · S. Shenkin
Department of Medicine for the
Elderly, The University of Edinburgh,
Edinburgh, UK

Abstract Interpretation of brain images from older patients requires knowledge of changes that occur with healthy ageing. We constructed and tested a reference template for older subjects. We used MR images from normal subjects aged 65–70 and 75–80 to generate average age-specific images. We ranked the T2-weighted images by worsening brain tissue loss to create a diagram of key centiles. Two neuroradiologists tested the template during routine reporting; eight radiologists read 99 MR examinations without and then with the template. Fifty-four subjects (65–70 years) and 25 subjects (75–80 years) formed the reference images. For the two neuroradiologists, the reference template reduced the abnormal scan reporting from 28/42 without to 3/42 with the template. Of 99 MR examinations assessed by eight radiologists, 39/58 scans (67%) reported as moderate or severe atrophy without the template were reported as normal with the template (p=0.00011). Reference templates of the brain at older ages can "calibrate" radiology reporting. They could also be useful for research into ageing and related conditions. Larger numbers of examinations from more diverse populations and at different ages are required to increase the versatility of these templates.

Keywords Ageing · Cerebral atrophy · Magnetic resonance imaging · Normative data · MRI · Normal brain · Atrophy

Introduction

Loss of brain parenchyma with advancing age is well recognised. Physiological variation between individuals increases significantly with age, creating a wider spread of "normality" at extremes of age [1]. Dementias are associated with more rapid volume loss on sequential scans than the natural volume loss of normal ageing [2]. However, during routine radiological reporting, it is common to be faced with CT or MR examinations from only one timepoint, and there are no reference images of the brain for normal older populations. Therefore, it can be difficult to determine whether the brain images of an older patient are normal or whether there is more than average brain volume loss compared with normal populations of the same age.

A set of images indicating the median and outer limits of the normal range by deciles or quintiles of increasing age

would be useful for several reasons. With reference images, radiologists could rapidly report not just whether an individual's cranial cross-sectional examination was within normal limits, but at what point within the normal range (e.g. "on the 25th centile for age") by visual comparison. This might improve consistency of reporting between radiologists and enable quantification of change over time without need for computational analysis. Many types of neuroimaging research require image registration to a standard brain template originally produced from healthy subjects aged 20–35 [3]. Templates derived from normal older subjects might be better for research on older subjects.

Ongoing studies are acquiring brain images from different-aged subjects, but although it is possible to produce high quality averages [4], there are few age-specific templates [5] and none as yet for older subjects. To be included in such templates, it is crucial that the subjects

are proved to be healthy and have normal cognitive function. Therefore demographic details, medical histories, examination and laboratory tests are very important. We developed and performed initial testing of two age-specific normal reference templates for older people.

Methods

Subjects We used T1- and T2-weighted MR images from two cohorts acquired in studies of normal ageing. The "Normal Ageing Brain" Study [6] included normal men aged 65 to 70 years (n=97). The "Simpson's study" [7] included community-dwelling men and women aged 75 to 80 years (n=105). Both studies had local ethical board approval, and all participants gave written informed consent.

Both cohorts underwent detailed assessment of past and current medical and psychiatric history, examination, full laboratory testing and neuropsychological tests of prior and current mental ability [6, 7]. Any non-healthy subjects were excluded. We also excluded subjects with blood pressure over 150 systolic or 90 diastolic at the time of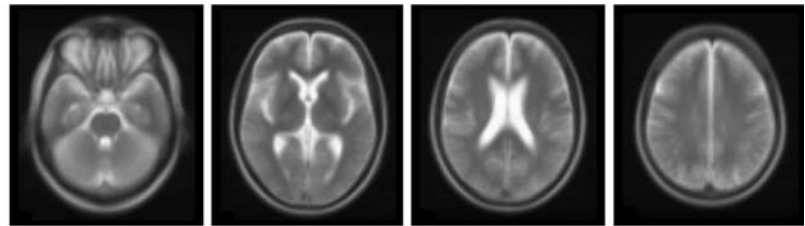 initial testing even though they were not receiving antihypertensive treatment and had no previous diagnosis of hypertension.

Image processing Details of the image acquisition sequences (T2-weighted axial and T1-weighted volumetric imaging) have been given previously [6, 7]. We spatially co-aligned images to a standard template and corrected for global differences in brain size and shape using a 3D affine transformation [8]. We aligned the T1-weighted images to the standard Montreal Neurological Index-152 brain template and the T2-weighted images to the Montreal Neurological Index-compatible T2 template (in SPM99; www.fil.ion.ucl.ac.uk/spm), sub-sampled to give 6-mm-thick slices more appropriate for the T2 images acquired in this study. We normalised the image intensity to ensure that all images had similar intensity ranges and contributed an equal weighting in the image averaging process. We calculated the mean image intensity on a voxel-by-voxel basis for each of the two cohorts to produce their respective average brain images (Fig. 1).
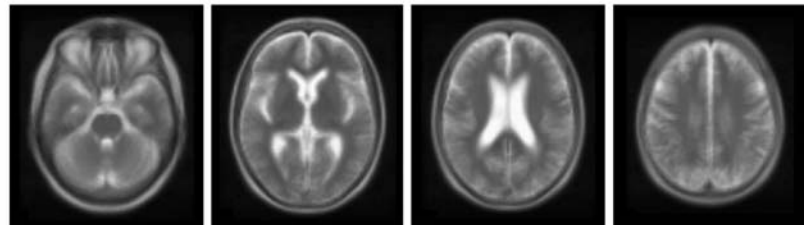
A neuroradiologist ranked the T2-weighted images, for the 65–70 and 75–80 year olds separately, from the smallest to largest ventricles and sulci. The MR examinations that represented the smallest (i.e. equivalent to 5th

Fig. 1 Average brain templates obtained from the two different age groups showing four representative slices from the transverse T1- and T2-weighted templates
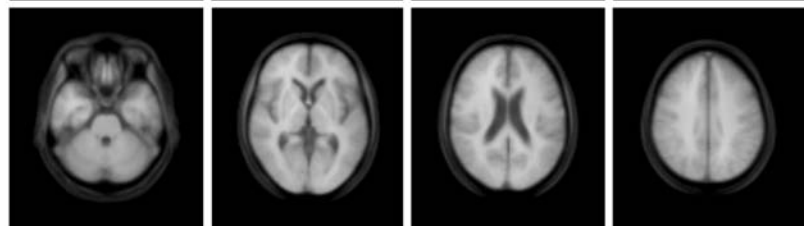
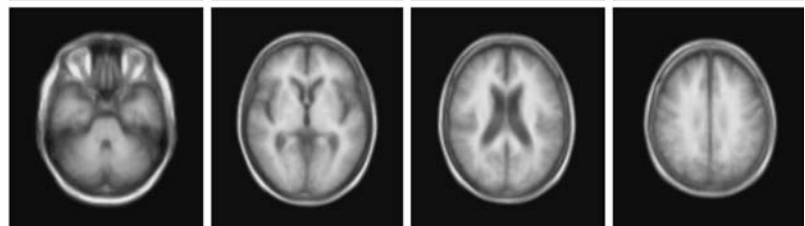centile), largest (95th centile), median and 25th and 75th centiles from each cohort were selected to represent the range of normal cranial appearances. Standard axial images from the upper brainstem, third ventricle, bodies of lateral ventricles and centrum semiovale were displayed on a template with the averaged brain slices—the reference template (Fig. 2).

Testing of template We tested the usefulness of the reference template in two ways. Two neuroradiologists used the reference templates to guide their interpretation during routine radiological reporting of a mixture of MR and CT brain images. The images were as encountered during standard reporting sessions and included patients aged over 60 with a range of indications. The neuroradiologists reported whether the degree of brain tissue loss

was within normal limits or excessive for age, firstly without referring to the reference template, and then with the reference template and any change in the radiologists' interpretation was noted.

Eight radiologists (some trainees in the final 3 years of general radiology training and some consultant neuroradiologists) read T2-weighted and FLAIR MR examinations from existing research studies of older people with minor stroke. The eight radiologists categorised the degree of atrophy in four categories—none, mild, moderate, severe—as though reporting the MR examinations routinely, from hard copies. Each radiologist examined a different set of examinations. After at least 2 weeks, the radiologists examined the same examinations again, this time using the reference template, and categorised the brain appearance to the nearest centile (5th, 25th, 50th, 75th, 95th), or if more atrophied than the
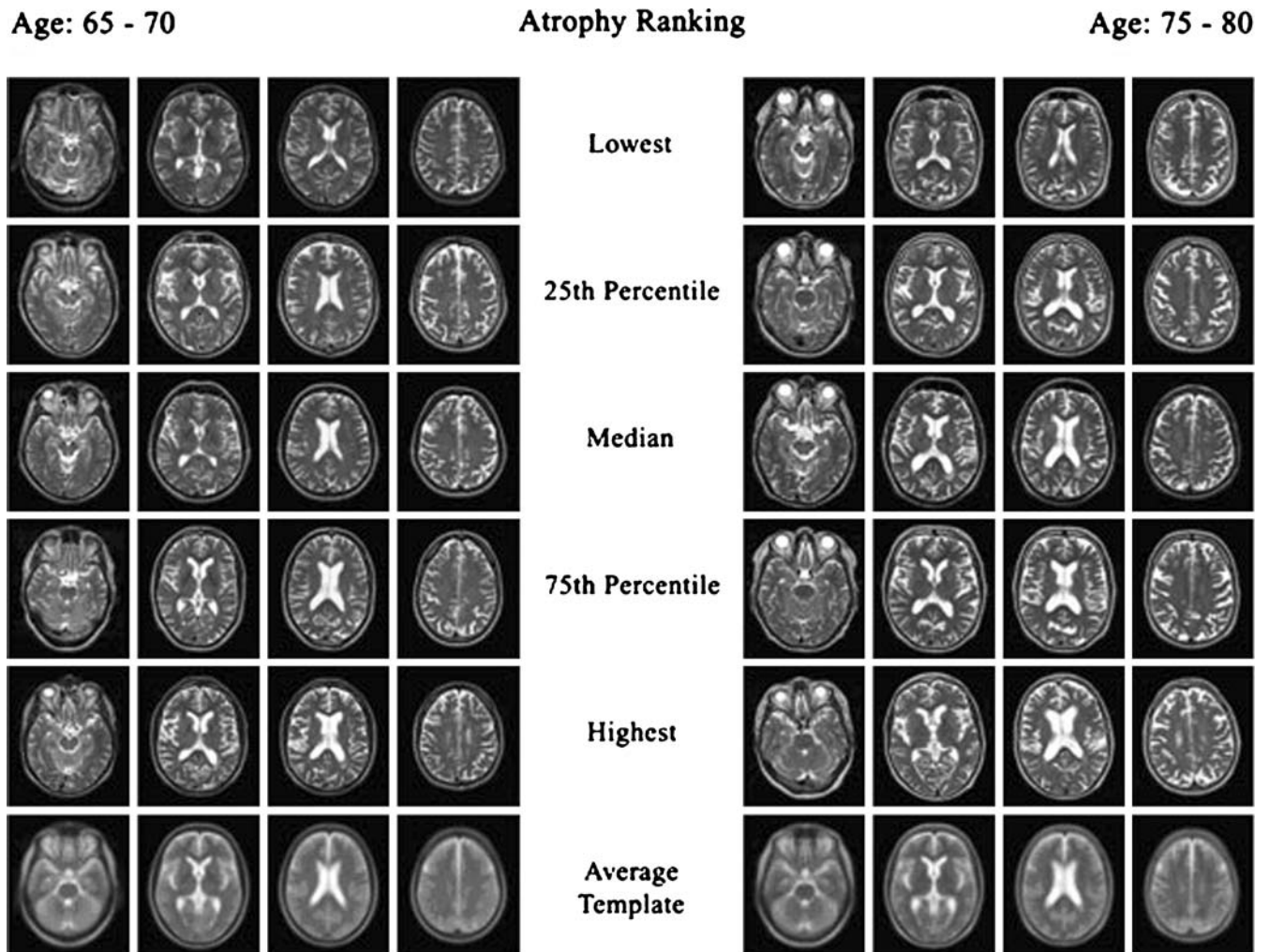

Fig. 2 T2-weighted transverse sections in order of least to most volume loss for the two age groups studied. Images shown are representative of the lowest, 25th percentile, median, 75th percentile and highest degrees of atrophy, along with the average template (bottom) obtained from the automated averaging process

95th centile, into two further categories ("worse than" and "much worse than" the 95th centile). The radiologists used the reference template closest to the patient's age.

Analysis We compared the proportion of examinations reported as normal or abnormally atrophied by the two neuroradiologists using chi-squared tests. We plotted the eight radiologists' readings without and with the reference template using all categories. To compare the effect of the template on judging the MR examinations to show "normal" versus "abnormal" appearance, we condensed the categories: without the template, "none" and "mild" atrophy as "normal" and "moderate" and "severe" atrophy as "abnormal" with the template, "5th" to "95th" centiles as "within normal limits" and ">95th" and ">>95th" centiles as "atrophy more than normal for age". Using this condensed classification, we compared all trainees combined, all consultants combined and all radiologists combined, using Fisher's exact test. We then compared the shift between individual categories of atrophy, without and with the template, for the trainees combined, consultants combined and all radiologists combined using the Bhapkar modification of the chi-squared test (Uebersax JS. User guide for the MH program vers. 1.1 *Statistical Methods for Rater Agreement* website. 2006. http://ourworld.compuserve.com/homepages/jsuebersax/mh.htm. Accessed: *01/08/2007*). This tests marginal homogeneities, ie whether individual radiologists increased or decreased their assessment of brain size when using the template, compared to without the template. This comparison requires the same number of categories in each classification, so we condensed the categories with the template so as to have the same number of categories as without the template: "5th" and "25th" centiles = 1; "50th" and "75th" = 2; "95th"and ">severe" = 3 and ">>severe" = 4.

Results

The final T2- and T1-weighted reference templates for 65 to 70 year olds (average age 67, 54 subjects) and for 75 to 80 year olds (7 male and 18 female, average age 77, 25 subjects) included only half and less than a quarter of the total subjects from their respective studies. We had excluded about half of the 65–70-year-old cohort because their blood pressure was elevated at initial testing and >75% of the 75–80 year olds due to concurrent medical condition (mostly hypertension or diabetes) or past medical history. This was in addition to any exclusions at recruitment into the primary cohorts due to finding incidental abnormalities shown by MR (e.g. meningioma, pituitary tumour, arachnoid cyst, etc.). Therefore, the ratio of subjects recruited in the primary studies to those included in the reference templates was about 2:1 at age 65–70 and 4:1 at 75–80.

Visual inspection of Figs. 1 and 2 shows a modest although clear increase in ventricular and sulcal size between ages 65 to 70 and 75 to 80. Also, displaying the four standard slices from the subjects on the 5th to 95th centiles for atrophy (Fig. 2) helped to overcome the "blurred" margins of the average images. The range of brain volume loss is wide for both age groups, with the samples representing the 75% and 95th centiles for the 65–70 year olds overlapping with up to the 50th centile of the 75–80 years olds (Figs. 1 and 2). The range of normal brain appearances appears wider at 75–80 years.

In routine reporting, the two neuroradiologists had rated 42 examinations in patients aged 67 to 90 years, median 79 years. Without referring to the template, 14 were said to be abnormally atrophied for age (33%), 14 were said to show a normal degree of atrophy for age, and 14 (33%) were interpreted as being less atrophied than would be expected for age. After referring to the template, only three examinations (6%) were reported as outside the normal range for age (i.e. more atrophied than the 95th centile), the rest all being within the normal range (p=0.001).

The eight radiologists categorised 99 examinations in total on two separate occasions. Each individual radiologist viewed between 7 and 22 MR examinations. Without the template, 58/99 MR examinations were classed as showing moderate or severe atrophy, but with the template 39/58 (67%) were classed as within the 5th to 95th centiles for age (Fig. 3). We considered that "moderate" or "severe" atrophy without the template was equivalent to "atrophy more than normal for age" with the template; "none" or "mild" without the template was equivalent to "within normal limits" (i.e. 5th to 95th centile) with the template (Table 1). Of 80/99 examinations classed as "within normal limits" with the template, 39/80 had been classed as "moderate or severe atrophy" without the template (2p=0.000011, Fisher's exact test). The results were very similar for trainees and consultants, with consultants reporting 38.9% (p=0.03) and registrars 40.0% (p=0.00069) of the scans as abnormal (moderate or severe atrophy) without the template, but
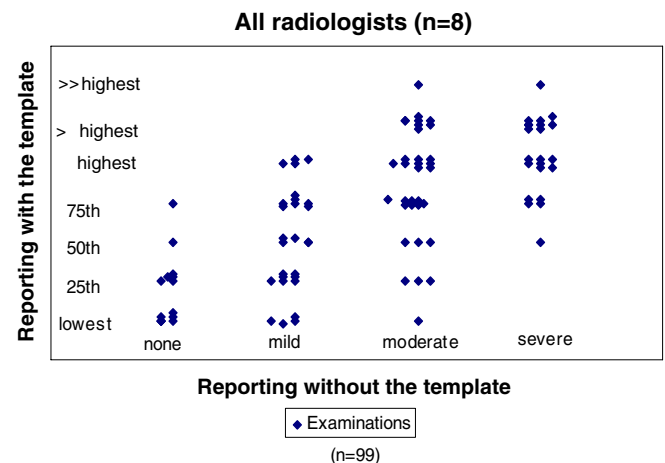

Fig. 3 Comparison of classification of 99 MR examinations by eight radiologists, with versus without the age-specific reference template

Table 1 Effect of older brain template on number of MR examinations reported as normal or abnormal by eight radiologists

With template	Registrars (n=63)		Consultants (n=36)		Total of all radiologists (n=99)	
	Without template					
	Normal (none/mild atrophy)	Abnormal (moderate/severe atrophy)	Normal (none/mild atrophy)	Abnormal (moderate/severe atrophy)	Normal (none/mild atrophy)	Abnormal (moderate/severe atrophy)
Normal (lowest-highest centiles)	26	**25 (40%)**	15	**14 (38.3%)**	41	**39 (39.4%)**
Abnormal (>highest, >>highest centiles)	0	12	0	7	0	19
Difference with use of template (two-tail p value, Fisher's exact test)	p=0.00069		p=0.03		p=0.000011	

"within normal limits" (5th–95th centiles) with the template (Table 1). The Bhapkar test of marginal homogeneity (Table 2) showed that, when a more detailed categorisation was retained, the classification without the template was significantly different to that with the template for trainees alone, consultants alone and all radiologists together (p= 0.0006 or less).

Discussion

We developed and tested a reference template of brain images of normal older subjects. The template reduced substantially the number of examinations reported as being outside the normal range for age. While further development is needed to increase the number of brains contributing to the template, from the same as well as other ages and different populations, this raises points for further debate. What is normal? Can MRI or CT expressing a broad range of ventricular/sulcal size be combined, or does the increasing physiological variation with age mitigate against "averaging"? How many subjects might need to be assessed to identify truly normal individuals?

We used MR examinations obtained from healthy older people in one city to produce these templates. We were extremely careful to make sure that the subjects were physically and cognitively normal. Subjects were not included simply because their brain images were considered normal. Subjects with cardiovascular disease, hypertension, diabetes and mild cognitive impairment (which may only become apparent when specifically sought) cannot be considered normal and should be excluded, as these vascular risk factors are associated with cerebral white matter changes and atrophy [2]. The performance of such careful screening was labour intensive and expensive. We excluded almost half of the cohort aged 65–70 and three quarters of the cohort aged 75–80 who had already been through detailed physical and cognitive testing to be included in the primary studies because of vascular risk factors. We were fortunate that the primary studies were funded at a level that enabled such detailed clinical assessment.

Radiologists' familiarity with the range of normality in their local population may be rather subjective and may result in patients being reported as abnormal when their degree of brain volume loss falls within that found in community-dwelling healthy normal older people, as happened with about half the examinations rated by the eight radiologists. Reference images like these could bring much-needed objectivity to the interpretation of brain imaging of older patients in routine reporting as well as for image normalisation in imaging research. For example, in a patient of 67 years, it was possible to say "the patient's brain appearance is on the 25th centile for a 65–70 year

Table 2 Difference in MR reporting of brain atrophy between normal and abnormal categories by eight radiologists without versus with the older brain template

		Registrars				Consultants				All radiologists			
		With template											
		1	2	3	4	1	2	3	4	1	2	3	4
Without template	None	5	1	0	0	6	1	0	0	11	2	0	0
	Mild	8	10	2	0	3	3	2	0	11	13	4	0
	Moderate	2	8	13	1	2	6	5	0	4	14	18	1
	Severe	0	2	11	0	0	2	5	1	0	4	16	1

Bhapkar test of marginal homogeneities. Registrars: p=0.0000; consultants: p=0.0006; all radiologists: p=0.0000. 1=5th–25th centiles; 2= 50th–75th centiles; 3=95th centile and >severe; 4 = >>severe atrophy

old". Or, in a patient aged 80 years: "the patient's brain appears well-preserved, the appearance being consistent with a normal brain at age 65 years or less." As the range of normality increases with age, a chart of normal age-specific brain scans in radiology reporting rooms might help maintain consistency of reporting, keeping the radiologist "calibrated" to the correct range of brain volume changes.

The image processing procedure used in this work is relatively simple. The global affine transformation accounts for general differences in brain size and shape. More sophisticated alignment techniques have been described, such as those utilising non-linear image registration [4]. However, it is unclear whether these methods are any more reliable at producing an 'average' alignment due to the difficulty in constraining the realignment in a realistic manner, an extremely challenging problem. The current scaling procedure is effective for images that exhibit similar contrast, such as those originating from the same MR machine using the same sequence, but more sophisticated intensity normalisation methods may be required to combine images originating from different MR machines and different imaging sequences. It is important to find methods for overcoming differences between machines. Improved methods for averaging brains are needed to account for the variation in cortical sulcal patterns at any age, but particularly for older people where some features of atrophy may be masked by current averaging techniques. This is a general problem in studies of ageing-associated changes on MRI. Techniques such as "boot strapping" or other statistical methods (see http://brainvisa.info) may help preserve individual detail during the averaging process, but require further evaluation.

The study has limitations. These patients were all extensively cognitively tested at the time of their scanning and were within normal limits. However, we do not know whether any of the patients developed cognitive impairment subsequently. While rapidity of atrophy is associated with development of dementia, we do not know if patients with brain appearance greater than the 50th centile are at higher risk of cognitive impairment (or how quickly) than those whose MR findings fall below the 50th centile for age. It would have been desirable for a far larger number of radiologists to look at a larger number of examinations to confirm these findings. The reference template could be improved by addition of more examples than just the five centiles and the average. Some of the eight radiologists commented that there were too many images on the template. Sometimes the sulci seemed disproportionate to the ventricles and vice versa, making choosing the closest match from the template difficult.

We do not know if these standard images are relevant to other geographical populations where ethnicity, nutrition, education, urban or rural dwelling could all influence brain appearance at older ages. Radiologists and imaging scientists should consider testing this template in their own population prior to routine clinical use, as well as developing their own locally relevant templates and contributing to national or international cohorts. Creation of further templates from subjects in different age bands, populations and MR machines should be encouraged. Projects are underway to produce normal brain atlases (e.g. www.nbirn.net/ and www.cma.mgh.harvard.edu/ibvd), and a paediatric normative brain has been produced [5]. To avoid creating misleading reference images, subjects contributing must be carefully screened for disease and cognitively assessed. These should not just be "normal" images from patients referred to imaging to investigate specific symptoms or signs, as these patients may not be normal even if the brain images are unremarkable.

References

1. Abbott A (2003) A new atlas of the brain. Nature 424:249–250
2. Fox NC, Freeborough PA, Rossor MN (1996) Visualisation and rates of atrophy in Alzheimer's disease. Lancet 348:94–97
3. Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM (1993) 3D statistical neuroanatomical models from 305 MRI volumes. Proc IEEE Nuclear Science Symposium and Medical Imaging Conference 1813–1817
4. Shen D, Davatzikos C (2003) Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. NeuroImage 18:28–41

5. Wilke M, Schmithorst VJ, Holland SK (2003) Normative pediatric brain data for spatial normalization and segmentation differs from standard adult data. Mag Res Med 50:749–757

6. MacLullich AMJ, Ferguson KJ, Deary IJ, Seckl JR, Starr JM, Wardlaw JM (2002) Intracranial capacity and brain volumes are associated with cognition in healthy elderly men. Neurology 59:169–174

7. Shenkin SD, Bastin ME, MacGillivray TJ, Deary IJ, Starr JM, Wardlaw JM (2003) Childhood and current cognitive function in healthy 80-year-olds: a DT-MRI study. Neuroreport 14(3):345–349

8. Jenkinson M, Smith S (2001) A global optimisation method for robust affine registration of brain images. Med Image Anal 5(2):143–156