Serial brain MRI at 3–6 month intervals as a surrogate marker for Alzheimer's disease

[1]K M BRADLEY, MRCP, FRCR, [2]G M BYDDER, FRCR, [1]M M BUDGE, FRACP, [2]J V HAJNAL, PhD, [2]S J WHITE, DCR, [3]B D RIPLEY, PhD and [1]A D SMITH, DPhil

[1]*Oxford Project to Investigate Memory and Ageing, University of Oxford, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE,* [2]*Magnetic Resonance Unit, MRC Clinical Sciences Centre, Hammersmith Hospital, London and* [3]*Department of Statistics, University of Oxford, Oxford, UK*

Abstract. A surrogate marker is needed for Alzheimer's disease (AD) both to aid diagnosis and to assess interventions. Despite widespread use, brain imaging markers have largely been confounded by overlap with "normal" ageing. 39 elderly subjects completed up to four serial volumetric brain MRI scans with intervals from 2.5 months to 7 months. By National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) criteria, five subjects had probable AD, two possible AD and 32 were negative for AD, although this group included memory-impaired subjects. Total brain and ventricular volumes were measured for each scan, and rates of change for each interval calculated. The rate of change in ventricle-to-brain ratio was 15.6% per year (standard deviation (SD) 2.8%) for probable AD compared with 4.3% per year (SD 1.1%) for negative AD ($p<0.001$). There was no significant difference between these groups' mean ventricle-to-brain ratios measured at a single time point ($p=0.25$). Rates of change in brain or ventricular volume over time also differed between the two groups ($p<0.001$). Power calculations reveal that to detect a 20% reduction in the excess rate of atrophy of a probable AD cohort in just 6 months, with 90% power, 135 subjects would be required in each arm of a randomized placebo controlled trial. For a 30% reduction in the excess rate of atrophy, 61 subjects would be required. Rate of change analysis makes serial brain MRI a valuable surrogate marker for Alzheimer's disease. Since only 6 months or less is required between scans, this procedure has both clinical relevance and potential for monitoring interventions.

Based on recent advances in understanding the pathology of Alzheimer's disease (AD), many of the major pharmaceutical companies are developing drugs that may slow, or hopefully arrest, the pathological lesions in AD. Since these are likely to be disease modifying agents, to be of benefit they will need to be administered early in the course of the disease, before there is irreversible loss of significant numbers of neurones and synapses. This is the time when clinical diagnosis is most difficult. A surrogate marker which could provide quantification of disease progress would aid diagnosis and help identify which of any new agents are likely to be of most benefit.

Since the early days of CT in the 1970s, and subsequently with MRI, it has been observed that many patients with AD have relatively atrophic brains. However, early attempts to quantify this atrophy as a useful marker in individuals were largely frustrated by overlap with normal ageing [1]. With the realization that Alzheimer pathology typically starts in the entorhinal cortex before

involving the hippocampus, probably for several years, before spreading cortically [2], imaging research focused on measurement of medial temporal lobe structures [3–7]. Using serial CT scans it was shown that a simple linear measurement of the thickness of the medial temporal lobe provided an estimate of the rate of progression of atrophy, but only over a period of several years [4]. Unfortunately, the complex anatomy and indistinct borders of structures within the medial temporal lobe means that this structure has defied robust, automatic measurement from MRI scans. Skilled personnel have to undergo an extensive learning phase, analysis is time-consuming and the process is fraught by intra- and inter-observer variation. Even in the best institutions, intra- and inter-observer variation, when quoted, is usually ⩾5%.

Recently, with the availability of powerful computer hardware, subvoxel image registration techniques have become possible. These methods more easily permit longitudinal assessment of change, not just cross-sectional analysis. Fox and colleagues, measuring initially in familial AD patients and then in sporadic AD patients,

demonstrated that the brains of AD patients atrophied at a rate significantly greater than age-matched controls on MRI scans taken approximately 1 year apart [8–11]. Their work, although in relatively young patients, points the way to an aid for monitoring disease progression and evaluation of potential treatments.

This paper reports results from serial brain MRI examinations of a heterogeneous elderly population including subjects with National Institute of Neurological and Communicative Disorders and Stroke—Alzheimer's Disease and Related Disorders Association (NINCDS) [12] probable and possible AD. Image registration and subtraction were used to visualize and identify patterns of brain changes. Total brain and ventricular volume were quantified directly from MRI data. The first aim of this study was to prospectively follow individual subjects repeatedly, every 3–6 months, to identify the minimum interval required to separate AD from typical ageing. The second aim was to determine the potential of serial MRI to provide an objective marker of disease progression.

Methods

Subjects

39 subjects who had volunteered for the Oxford Project To Investigate Memory and Ageing (OPTIMA), a prospective, longitudinal, clinico-pathological study of ageing both in "normal" elderly and memory-impaired subjects, completed serial MRI. All subjects gave informed consent and the research was approved by the Central Oxford Research Ethics Committee. All the volunteers in this study have given consent to future post-mortem examination. By NINCDS criteria, 5 subjects were classified as probable AD, 2 as possible AD and 32 as negative (Table 1). Age, mini mental stage examination (MMSE) and the cognitive score (CAMCOG) from the

Table 1. Subject characteristics

Alzheimer's disease state	Subject	Age (years)	MMSE (0–30)	CAMCOG (0–107)
Probable	1	64.7	15	62
	2	67.2	26	85
	3	59.6	14	51
	4	56.8	22	90
	5	60.2	6	24
Possible	1	77.4	24	81
	2	81.9	21	82
Negative	32	73.5	28.6	100
mean+(SD)		(17.1)	(1.8)	(4.8)

MMSE, mini mental stage examination score; CAMCOG, Cambridge Examination for Mental Disorders in the Elderly cognitive score; SD, standard deviation.

Cambridge Examination for Mental Disorders in the Elderly were taken at the time of the first MRI scan. A record was made of any change in medical history or medications taken. Apolipoprotein E (ApoE) genotype was determined by standard methodology.

It is important to emphasize that the "negative" subjects were a mixed elderly population, two actually referred with objective memory complaint. Six of the NINCDS negative subjects had changes on brain CT that are frequently reported as associated with vascular disease. Five subjects demonstrated periventricular white matter low attenuation. Two of these had further lesions within basal ganglia, one of whom had an old, small infarct involving the head of a caudate nucleus and adjacent anterior limb of internal capsule. The sixth subject, who subsequently died from a stroke, had subtle low attenuation involving both basal ganglia. One NINCDS negative subject, who had undergone two scans, was excluded due to the presence of an inexplicable artefact present in the soft tissues of the difference image. Her brain data was consistent with her category. Two subjects (one NINCDS negative and one NINCDS probable) refused subsequent scans after the baseline scan and a further NINCDS negative subject was not asked to return after her baseline scan since she had sustained a femoral shaft fracture.

Since the MRI was performed, four subjects have died. One NINCDS probable subject confirmed as definite AD by histology using Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria [13]. Three NINCDS negative subjects were all confirmed CERAD normal aged brain. One of these had small vessel disease (not evident on CT/MRI) and another died from a large middle cerebral artery infarction.

MRI acquisition

All MRI scans were performed using a 1.0T Marconi Medical Systems HPQ+ scanner (Picker International, Cleveland, OH) at the Hammersmith Hospital, London. Subjects were scanned up to four times, with intervals ranging from 2.5 months to 7 months. For each subject, all scans were performed at approximately the same time of day on the same day of the week using a standardized radiographic protocol. Image data was obtained from a quadrature transmit receive birdcage coil using a three-dimensional radio frequency spoiled T_1 weighted acquisition, employing non-selective excitation pulses (TR, 21 ms; TE, 6 ms; flip angle, 35°). Images were acquired in the sagittal plane with a head-to-foot frequency encoding direction to avoid aliasing. The image matrix was 256×
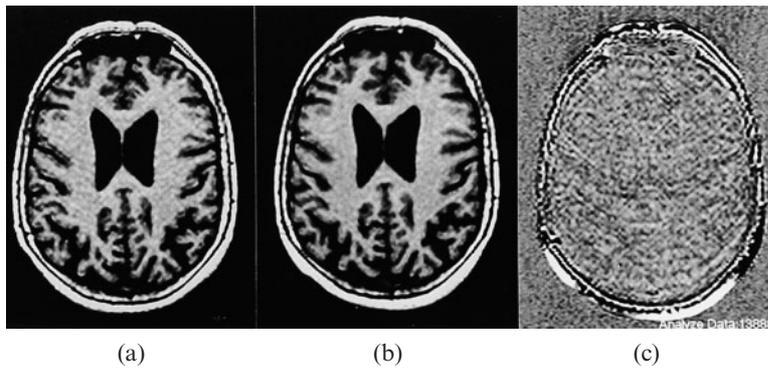
Figure 1. National Institute of Neurological and Communicative Disorders and Stroke Alzheimer's disease negative subject. (a) Baseline image; (b) registered image 6 months later; (c) registered difference image.

152×114 from a volume of $25 \text{ cm} \times 25 \text{ cm} \times 18 \text{ cm}$, and this was then zero filled to produce 256×256 matrix images with a section thickness of 1.3 mm. Machine calibration and performance were checked by acquisition of data from test phantoms at regular intervals throughout the course of the study.

MRI processing and interpretation

All MRI processing and interpretation was performed by investigators experienced in the techniques and blinded to any subject information. For each subject, the initial MRI scan was designated as a baseline and all subsequent scans were positionally registered to the baseline using a rigid body method based on the sum of squared voxel intensity differences [14]. Follow-up images were re-sampled onto the voxel matrix of the baseline images using sinc interpolation, and subtraction images were produced. Only the brain was used for co-registration, soft tissues were excluded. This is crucial since soft tissue position is variable, *e.g.* due to scalp positioning in the head holder, chin position, or swallowing, and these variations could compromise overall registration. The technique has been extensively verified with phantom studies and human studies with same day test-retest, as well as time course studies to monitor long-term machine and process stability. All images were also reformatted into the transverse plane with isotropic 1 mm^3 voxels.

Two complementary, but separate, methods of analysis were employed. Difference images were produced by subtracting registered images from the baseline, or subsequent images from each other, when there were more than two scans. These subtraction images cancel constant features but clearly demonstrate even subtle changes between images that could easily escape detection on the original scans. Features revealed by these difference images can usually be appreciated by hindsight on the original scans (Figures 1 and 2). This visual check was performed, blinded to the volume data, by a radiologist (GMB) experienced in the technique, to produce a visual rating of volume change and to ensure no misregistration artefacts had occurred or new incidental pathology was present.

Brain and ventricular volumes were quantified by image segmentation, based on intensity thresholding and contour following, using a semi-automatic method. Coil bias field effects were removed using a non-parametric approach. Then, rather than re-normalizing all images independently, which could introduce systematic errors, the baseline images were used as a bias field reference to which all subsequent images were harmonized [15].

To segment the brain and ventricles, contours enclosing the boundaries of the structures of interest were manually drawn on the baseline images. These contours were used to exclude unwanted regions of the images where other tissues of similar signal intensity to brain or cerebrospinal fluid (CSF) might have been present. A histogram analysis of the intensities within the bounding contour was then used to automatically set an intensity threshold, which in turn
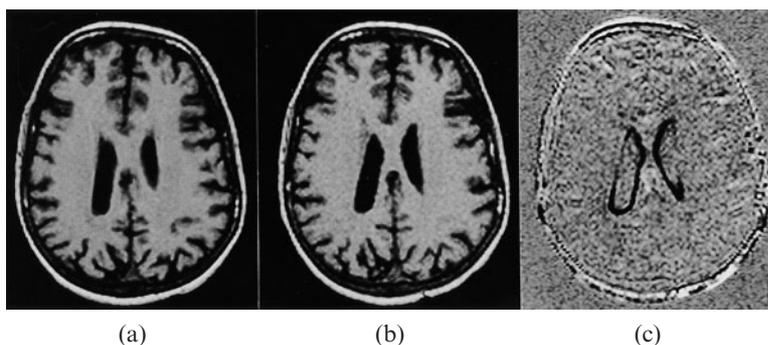

Figure 2. National Institute of Neurological and Communicative Disorders and Stroke Alzheimer's disease probable subject. (a) Baseline image; (b) registered imaged 6 months later; (c) registered difference image with black representing loss of tissue.

was used to identify and label all voxels associated with brain tissue or with ventricular CSF. Total brain (excluding CSF) and ventricular volumes were then calculated by multiplying the number of labelled voxels by the voxel volume in cubic millimetres. This method has been extensively validated on phantoms and normal control subjects.

Statistical analysis

A standard methodology for repeated measures studies was applied; that of linear mixed models. In this approach, the subjects are regarded as a sample from a population of individuals of that group, and the measure of interest (in this case brain and ventricular volume) follows a straight line over time for the duration of the study. This has a fixed component corresponding to the group to which a specific subject belongs, and a random component that gives the slope for a specific patient.

A second approach was also applied; that of generalized estimating equations. This considers the groups as a whole, and takes the repeated measurements of each subject as serially corre-lated. This is a less precise description of the situation and so is subject to greater uncertainty.

Software written by DM Bates and JC Pinheiro and incorporated in S-PLUS 2000 (Mathsoft DAPD, Seattle, WA) was used.

Results

The visual inspection of the subtraction images was consistent with the absolute measures of brain (B) and ventricular (V) volumes. The most conspicuous changes are observed at the brain/ventricle interface (Figures 1 and 2). Rates of change in volumes were calculated for each interval (range 2.5–7 months) for all subjects. These rates of change are expressed as percentage per year

Rate of change in ventricle: brain volume ratio

We calculated the rate of change of cere-bral ventricle:total brain volume ratio (V:B). By applying linear mixed effects models to compare second, third, and fourth scans with the baseline scan, all subject categories change over time, but the NINCDS probable group changes most rapidly (Figure 3). The NINCDS negative mean rate of V:B change was 4.3% per year (standard deviation (SD)=1.1%). The NINCDS probable mean rate of V:B change was 15.6% per year (SD=2.8%). The difference in rates of change is significant ($p<0.001$). There were only two NINCDS possible AD subjects who underwent

six scans in total. Interestingly, these two subjects' rate of percentage V:B change did not differ significantly from the NINCDS negative subjects, but the numbers are too small for more definite conclusions.

A generalized estimating equation analysis of this data gave similar results. The NINCDS negative mean rate of V:B change was 5.6% per year (SD=0.7%). The NINCDS probable mean rate of V:B change was 17.9% per year (SD=0.2%). Again, the difference is significant ($p<0.001$).

Change in brain volume over time

The NINCDS negative and possible groups have brains that shrank approximately 0.2% (SD=0.23%) per year. The NINCDS probable group's brains shrank 2.14% (SD=0.52%) per year. This difference is significant ($p<0.001$).

Change in ventricular volume over time

This change appears large owing to calculating percentage change, which permits comparison, rather than absolute volumes. The NINCDS negative group show a significant increase in ventricular volume of 4.1% (SD=0.9%) per year. In the NINCDS probable group the ventricular volume increased 13.0% (SD=2.4%) per year. The difference is significant ($p<0.001$). Both the changes in ventricular volume over time and brain volume over time are consistent with the rate of change of V:B.

Influence of age on ventricular volume

For this analysis, ventricular volume was divided by brain volume, to correct for head size, and only the initial scan of the NINCDS negative group was considered. Analysis was per-formed on a log scale, which is appropriate for the assessment of proportional change (Figure 4).

Large variations between subjects are revealed, but for the whole NINCDS negative group median values of V:B were approximately 4% (range 1.0–7.8%). There is a clear effect of age, with older patients having larger ventricles. The effect is approximately 3.8% (SD=1%) of ventricular volume per year of age in this popu-lation ($p<0.001$). This equates to an approximate doubling of ventricular volume in 20 years.

Influence of vascular changes and ApoE genotype

None of the six subjects with vascular changes on brain CT/MRI differed in their brain or ventricular volume changes over time, compared
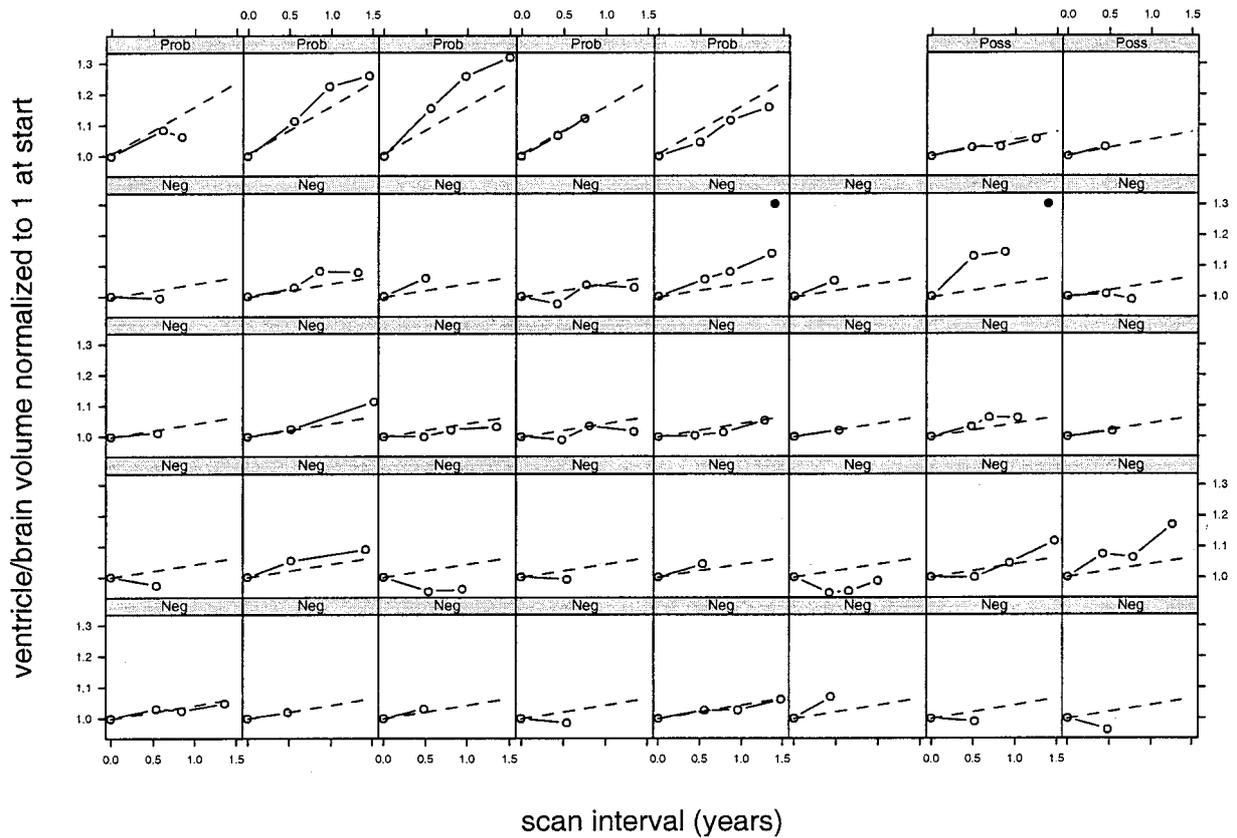
Figure 3. Rate of change of ventricular brain volume. MRI scans were performed at 2.5–7 month intervals, as shown by open circles for each subject. (– – –), calculated mean rate of change for each group: National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) probable (Prob), possible (Poss) and negative (Neg) for Alzheimer's disease; (●), two NINCDS Neg subjects referred with objective memory complaints.

with the rest of the NINCDS negative group. ApoE4 status did not significantly alter the rate of brain changes in this cohort ($p=0.52$).
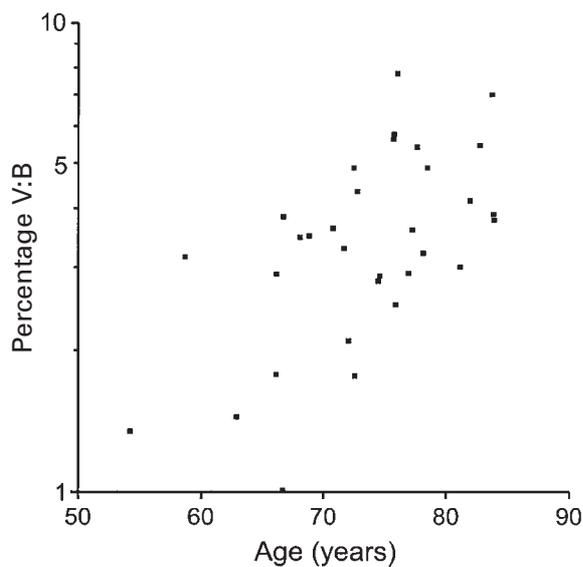

Figure 4. Influence of age on brain ventricular volume. V:B, cerebral ventricle:total brain volume ratio.

Discussion

This study demonstrates how MRI can distinguish subjects with probable AD from a mixed elderly population over a period of months. These differences are accurately quantified so the impact of any intervention may be directly assessed. Although attention to detail is required for consistency of scan acquisition, in order to make subsequent processing as straightforward as possible the measures described require no detailed neuroanatomical knowledge. Most of the process is semi-automated, requiring only supervisory checks.

The production of subtraction images can potentially be fully automated and, although currently performed on workstations, given enough demand manufacturers would be able to introduce appropriate software directly onto MRI consoles. Images would then be available for inclusion in a standard radiological report. This would make the technique practical even in busy general hospitals, with the added advantage that misregistration artefacts and confounding pathology are easily appreciated. Since this technique shows clear differences in only a few months, and

most new referrals attend out-patients at least twice, there is scope for at least visual assessment of subtraction images to be included in patient assessment. Subtraction images would almost certainly also be of diagnostic help in fronto-temporal dementia, with its characteristic patterns of atrophy.

The method of analysis used here, quantifying rate of change of ventricular volume, bridges the gap between those who measure medial temporal lobe structures, (most often the hippocampus) and those who measure longitudinal change in whole brain. Measuring the rate of change of ventricular volume seems logical after assessing the difference images, and expressing the ventricular volume as a ratio divided by total brain volume as a ratio divided by total brain volume corrects for between-subject variation in head size. The hippocampus bulges up into the floor of the temporal horn of the lateral ventricle, so measurement of ventricular volume has direct relevance to medial temporal lobe atrophy. Furthermore, loss of brain substance, *e.g.* in medial temporal lobe, may cause ventricular enlargement that is not localized to the immediate vicinity of the atrophied brain. It is noteworthy that earlier CT studies have also shown a rapid increase in the volume occupied by the ventricles in subjects with AD [16–19], and DeCarli et al [19] suggested that the rate of dilatation of the ventricle was a better discriminator between AD and normal ageing than absolute size.

Our study demonstrates the problem undermining the use of measurements of brain structures to classify subjects after a single scan. If the percentage V:B at time zero of the NINCDS probable group is compared with the NINCDS negative group, there is no statistically significant difference ($p=0.25$), compared with the highly significant difference for the rate of change in the percentage V:B ($p<0.001$). The explanation for this is found within the graph of age *vs* ventricular volume for the NINCDS negative group (Figure 4), with baseline V:B ranging from 1.0% to 7.8%. It is this large variability in structural appearance of normal brains that has frustrated many investigators. The effect of age on ventricular volume agrees with the established literature [20].

The 32 NINCDS negative subjects in the study were far from the typical selected controls of most studies, which emphasizes the power of this technique. Our expectation was that more of the negative group would have shown rapid atrophy since, given their age, it is reasonable to expect this group to be contaminated, with some in the presymptomatic stage of AD. It will be interesting to see how the two subjects in the negative group, whose rate of change of V:B is

noticeably above the group mean, fare over the next few years. These were the only subjects within the NINCDS negative group referred with memory problems. 1 year after their final MRI scan, one of these two subjects has become classified as NINCDS possible, and is the only person in the cohort to have changed category as yet.

Our approach to quantification allows us to make power calculations for assessing potential treatments. A caveat for all these calculations is that there is no knowledge of the variability of brain changes in the treated group, since there has never been such a group. Therefore, an assumption is required that variability does not differ by group. In this study the difference between two measurements on any subject had a SD of 1.6% and subjects had a SD in their rates of change of V:B of 5.5%. The calculations are all to a power of 90%, to detect an α percentage reduction in the difference of rate of change of V:B between a treated and untreated (placebo) group of NINCDS probable AD subjects towards the rate of change of V:B for NINCDS negative subjects (Figure 5).

The reason that reducing the interval from 12 months to 6 months leads to only a modest increase in treatment group size is that the major variability is between subjects and not between measurements. This re-emphasizes the robustness of the measurements. Power calculations rely upon variance, which is very sensitive to
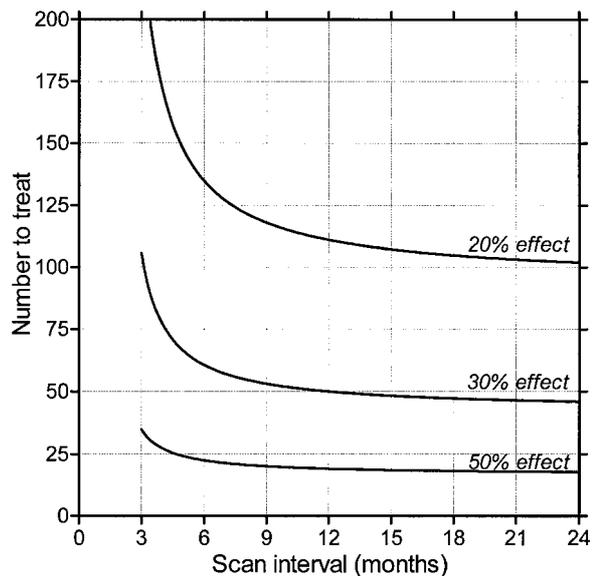

Figure 5. Power calculation (90%) to detect a reduction in the rate of change of cerebral ventricle:total brain volume between treated and untreated groups of National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) Alzheimer's disease (AD) probable subjects, compared with NINCDS AD negative subjects.

outliers. Our NINCDS negative group is likely to be contaminated with incipient AD, and if either of the two subjects with the greatest rate of change in this group is removed then the group size required for a 12 month interval, $\alpha=20\%$, falls from 110 to 88.

Our results for whole brain atrophy over time, probable AD=2.14% per year (0.52), negative AD=0.2% per year (0.23) are comparable to the brain changes described by Fox et al [11], who found probable AD=2.37% per year (1.11), controls=0.41% per year (0.47). If power calculations are performed with these figures, group size predictions are very similar. By incorporating ventricular volume change over time, percentage V:B numbers are reduced by approximately one fifth. We strongly believe that measurements and calculations of this kind should not be performed without the concomitant production of subtraction images to provide an internal consistency check.

The ability to quantify accurately and monitor brain changes is of significance to the potential design of drug trials. Clearly, the overall design of trials will be based on clinical and cognitive measures, but this imaging-based technique may play a useful role in rapid dose-finding or drug comparison studies. Also, since it is possible that disease modifying agents may only demonstrate delayed clinical benefit, it could be useful in the identification of drugs that warrant prolonged clinical trials.

Furthermore, with the inexorable advances in computing hardware and software combined with falling costs, a technique such as this that potentially needs no detailed neuro-anatomical knowledge may not be all that far away from general availability. This, in turn, could help achieve standardization of scan assessment between different institutions, as well as benefit a far greater proportion of the elderly population.

Acknowledgments

References

1. DeCarli C, Kaye JA, Horwitz B, Rapoport SI. Critical analysis of the use of computer-assisted transverse axial tomography to study human brain in aging and dementia of the Alzheimer type. Neurology 1990;40:872–83.

2. Braak H, Braak E, Bohl J. Staging of Alzheimer-related cortical destruction. Eur Neurol 1993;33: 403–8.

3. Jobst KA, Smith AD, Szatmari M, Molyneux A, Esiri ME, King E, et al. Detection in life of confirmed Alzheimer's disease using a simple measurement of medial temporal lobe atrophy by computed tomography [see comments]. Lancet 1992;340:1179–83.

4. Jobst KA, Smith AD, Szatmari M, Esiri MM, Jaskowski A, Hindley N, et al. Rapidly progressing atrophy of medial temporal lobe in Alzheimer's disease. Lancet 1994;343:829–30.

5. Jack CR Jr, Petersen RC, O'Brien PC, Tangalos EG. MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease. Neurology 1992;42: 183–8.

6. Jack CR Jr, Petersen RC, Xu YC, Waring SC, O'Brien PC, Tangalos EG, et al. Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease [see comments]. Neurology 1997;49:786–94.

7. Fox NC, Warrington EK, Freeborough PA, Hartikainen P, Kennedy AM, Stevens JM, et al. Presymptomatic hippocampal atrophy in Alzheimer's disease. A longitudinal MRI study. Brain 1996;119:2001–7.

8. Fox NC, Freeborough PA, Rossor MN. Visualisation and quantification of rates of atrophy in Alzheimer's disease [see comments]. Lancet 1996;348:94–7.

9. Rossor MN, Fox NC, Freeborough PA, Roques PK. Slowing the progression of Alzheimer disease: monitoring progression. Alzheimer Dis Assoc Disord 1997;11 (Suppl. 5):S6–9.

10. Fox NC, Warrington EK, Rossor MN. Serial magnetic resonance imaging of cerebral atrophy in preclinical Alzheimer's disease (letter). Lancet 1999;353:2125.

11. Fox NC, Cousens S, Scahill R, Harvey RJ, Rossor MN. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects [see comments]. Arch Neurol 2000;57: 339–44.

12. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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 1984;34: 939–44.

13. Mirra SS, Heyman A, McKeel D, Sumi SM, Crain BJ, Brownlee LM, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology 1991;41:479–86.

14. Hajnal JV, Saeed N, Soar EJ, Oatridge A, Young IR, Bydder GM. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. J Comput Assist Tomogr 1995;19:289–96.

15. Saeed N, Puri BK, Oatridge A, Hajnal JV, Young IR. Two methods for semi-automated quantification of changes in ventricular volume and their use in schizophrenia. Magn Reson Imaging 1998;16: 1237–47.

16. Luxenberg JS, Haxby JV, Creasey H, Sundaram M, Rapoport SI. Rate of ventricular enlargement in dementia of the Alzheimer type correlates with rate of neuropsychological deterioration. Neurology 1987;37:1135–40.

17. de Leon MJ, George AE, Reisberg B, Ferris SH, Kluger A, Stylopulos LA, et al. Alzheimer's disease: longitudinal CT studies of ventricular change. AJR 1989;152:1257–62.

18. Burns A, Jacoby R, Levy R. Computed tomography in Alzheimer's disease: a longitudinal study. Biol Psychiatry 1991;29:383–90.

19. DeCarli C, Haxby JV, Gillette JA, Teichberg D, Rapoport SI, Schapiro MB. Longitudinal changes in lateral ventricular volume in patients with dementia of the Alzheimer type. Neurology 1992; 42:2029–36.

20. Coffey CE, Wilkinson WE, Parashos IA, Soady SA, Sullivan RJ, Patterson LJ, et al. Quantitative cerebral anatomy of the aging human brain: a cross-sectional study using magnetic resonance imaging. Neurology 1992;42:527–36.