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

Objective: To evaluate the rate of brain atrophy calculated from serial magnetic resonance imaging (MRI) registration as a surrogate marker of disease progression for use in clinical trials in Alzheimer disease (AD).

Methods: Eighteen patients with mild to moderate AD and 18 age-matched normal controls underwent 2 MRI brain scans separated by a 12-month interval. Each individual’s later scan was registered to their first scan, and the volume of cerebral tissue loss calculated directly from the registered and subtracted MRI scan pairs. The mean and SD of the rate of brain volume changes were used to estimate the sample sizes that would be needed in a clinical trial with a drug anticipated to modify disease progression by varying degrees. Comparable sample size estimates were performed with data for other methods of monitoring rates of brain atrophy, extracted from published papers.

Results: The mean (SD) rate of brain atrophy for the patients with AD was 2.37% (1.11%) per year, while in the control group it was 0.41% (0.47%) per year. Based on these figures, to have 90% power to detect a drug effect equivalent to a 20% reduction in the rate of atrophy, 207 patients would be needed in each treatment arm. This assumes a 1-year placebo-controlled trial with a 10% patient dropout rate, and that 10% of scan pairs are unusable.

Conclusion: Registration of serial MRI volume images provides a powerful method of quantification of brain atrophy that can be used to monitor progression of AD in clinical trials.

Arch Neurol. 2000;57:339-344

ALZHEIMER DISEASE (AD) is characterized by gradual cognitive decline and progressive cerebral atrophy. At autopsy, the histopathological disease hallmarks of amyloid plaques and neurofibrillary tangles are accompanied by widespread synaptic and neuronal loss. Atrophy is a macroscopic consequence of these cellular losses that can be visualized in life using computed tomography or magnetic resonance imaging (MRI).1

For editorial comment see page 312

A wide range of drugs is currently under development for treating AD.2 The most successful class of compounds to date has been the cholinesterase inhibitors. These agents, which include tacrine,3 donepezil, rivastigmine, and metrifonate,6,7 have all produced statistically significant symptomatic improvement over 6 to 12 months in randomized, placebo-controlled clinical trials. The clinical trial designs used, eg, placebo-controlled, parallel group studies and assessed efficacy using end-point measurements. With the cholinesterase inhibitors, patients on active treatment seem to remain unchanged, as a group, while those who receive placebo decline. This has been widely interpreted as a delay in symptomatic progression. A vital question that now needs to be addressed is what effect are these compounds having on the biological progression of the disease; ie, are they able to alter the rate of neurodegeneration? Addressing this issue in clinical trials is the next major issue in drug development for AD.8,9

Cholinesterase inhibitors as a class were originally thought to be a symptomatic treatment only, yet there is now a suggestion from these early symptomatic protocol trials, and from in vitro research on amyloid processing, that there may be a disease-modifying effect. Moreover, there are numerous other potential candidates for disease-slowing effects, including neuroprotective, hormonal, and anti-inflammatory...
SUBJECTS AND METHODS

SUBJECTS

The Dementia Research Group at The National Hospital for Neurology and Neurosurgery, London, England, is involved in a wide range of longitudinal neuroimaging studies and clinical drug trials in Alzheimer disease. Patients recruited into these longitudinal studies routinely have MRI brain scans at 12-month intervals. The diagnosis of Alzheimer disease is made according to National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer’s Disease and Related Disorders Association criteria, and disease severity is estimated using formal neuropsychometry and clinical assessment including the MMSE. All data are stored in a centralized database (Microsoft Access 95 and SQL Server 6.5; Microsoft Corp, Redmond, Wash, 1993, 1996).

A database search was performed to identify all patients with clinically diagnosed AD who had undergone 2 MRI scans with approximately a 12-month interval between scans, and who had mild to moderate dementia (MMSE score, 10-26) at the date of their first scan. The search provided 18 patients with AD, 2 familial and 16 sporadic, with a mean interval between scans of 11 months. None of the patients was known to be taking active medication. Age-matched control subjects were selected from our pool of normal volunteers. Demographic details are given in Table 1.

All subjects had given their consent both for the MRI and for participation in longitudinal research studies. All studies had the approval of the Local Research Ethics Committee.

BRAIN IMAGING

T1-weighted, volumetric MRI scans were performed on a 1.5-T Sigma unit (General Electric, Milwaukee, Wis) yielding 124 contiguous 1.5-mm coronal slices through the head with a 256 × 128 image matrix (acquisition parameters: time to repeat/echo time/number of excitations/FLIP ANGLE, 35/5/1/35). Axial dual-echo sequences (T2- and proton-density weighted) were also acquired.

IMAGE ANALYSIS

Magnetic resonance images from the volumetric sequences were transferred to a Sun workstation (Sun Microsystems Inc, Mountain View, Calif) where registration and quantification of atrophy were performed retrospectively in a randomized and blinded fashion. Only the brain is used for the registration procedure, and semi-automated 3-dimensional image analysis procedures were used for approximate delineation of the brain from the rest of the MRI (scalp, neck, etc). The registration algorithm is fully automated and uses the complex structure of the whole brain, considered as a volume rather than a set of slices, to determine the rotations and translations necessary to place 1 brain image set accurately on the other. Each individual’s follow-up scan was registered to their baseline scan. The change in cerebral volume was computed directly from the registered image sets using a validated technique for automatically calculating the amount that cerebral boundaries have changed between scans. Cerebral volume loss was expressed as a percentage of initial total brain volume and annualized to give a rate of global atrophy.

STATISTICAL ANALYSIS

Data were analyzed using Stata version 5.0 (Stata Corporation, College Station, Tex, 1997). Sample size requirements were estimated using the standard formula:

Sample size = \( (u + v)^2 \times (\sigma_u^2 + \sigma_v^2) / (\mu_u - \mu_v)^2 \),

where \( u = 1.28 \) to provide 90% power and \( v = 1.96 \) to test at the 5% level; \( \mu \) and \( \sigma \) are the mean and SDs of rates of atrophy in the treatment and placebo groups (assumes \( \mu_1 = \mu_2 \)).

The following modifications were included:

1. The difference in annual rates of tissue loss between patients with AD and healthy age-matched controls was taken to represent the maximum possible effect a treatment could have (ie, 100% impact). If this assumption is not made, the potential for a therapeutic effect may be overestimated. A reduction of 20% in progression was therefore considered to be equal to 20% of this difference rather than 20% of the total loss in patients with AD.

2. The sample size thus derived was increased by 10% to allow for losses to follow-up.

3. The sample size was then further increased assuming that about 10% of scan pairs could not be used (eg, too much movement on 1 of the scans).

All calculations were performed based on the requirement that a trial should have 90% power to detect the specified treatment effect when a 2-sided 5% level of significance is used.
might not be considered clinically relevant. A disease-
slowing effect would be difficult to show and
effected by various magnitudes of effect. We further compared these sample size estimates with those derived from other methods of measuring progressive brain atrophy using published rates of ventricular enlargement and hippocampal atrophy.

The AD and control groups were closely matched for age, sex, and interval between scans (Table 1). The mean (SD) rate of cerebral atrophy was 2.37% (1.11%) of brain volume per year for the AD group and 0.41% (0.47%) per year for the control group. The change in mean (SD) MMSE scores was from 19.6 (4.1) at baseline to 17.1 (6.3) at second scan in the patients with AD and 29.2 (1.0) to 29.2 (1.1) in the control group.

The results of the sample size calculations for the serial MRI registration technique are shown in Table 2. This table can be interpreted for a drug with an anticipated ability to reduce the rate of cerebral atrophy by 20% over 1 year as follows: To have 90% power to detect a drug effect, allowing 0.24% per year for normal aging, a 10% dropout rate, and assuming that 10% of scan pairs are unusable, 207 patients will be needed in each treatment arm.

The calculations presented above suggest that this method of measuring rates of cerebral atrophy may be feasible for use in clinical phase 2/3 drug trials. The calculations take into account subject attrition and technical problems with scan acquisition, yet still indicate a realistic sample size given the cost of including serial MRI in a large clinical trial. The therapeutic effects used in the examples given ranged from 10% to 50%. Less than a 10% disease-slowing effect would be difficult to show and might not be considered clinically relevant. A disease-slowing effect of more than 50% is unlikely to need a surrogate marker of progression because the clinical effects would be so dramatic.

Serial MRI registration is a comparatively new technique for measuring rates of cerebral atrophy. Although a large number of articles describe the cross-sectional features and correlates of neuroimaging changes and volumetric measures in AD, only a relatively small number of longitudinal structural imaging studies of appropriate design have been published with atrophy rates and their SDs reported. A number of studies where comparable data are available are summarized in Table 3. The paucity of longitudinal studies is surprising considering their importance.

Comparisons of the different techniques used to measure atrophy are difficult because the patient data sets are necessarily different. Based on the published rates of atrophy of the structures measured, sample size calculations were carried out using an identical method to that used for the MRI registration technique. These sample size estimates are summarized in Table 4.

Most studies focused on ventricular enlargement. Of those listed, only 2 studies used MRI. Kaye et al studied elderly patients who were initially recruited "pre-dementia" and observed to 12 months postdiagnosis. Rates of atrophy might therefore be expected to be higher if based only on the period postdiagnosis. Jack et al (1998) conducted a careful longitudinal study measuring hippocampal and temporal horn volumes from volumetric MRI in 24 patients with AD and 24 controls. The interval between scans was almost 2 years. Registration was used prior to manual segmentation of the regions of interest resulting in high levels of reproducibility. The rate of hippocampal atrophy in the patients with AD (4% per year) was only 2.5 times that of the control group (1.6% per year), and the relatively small difference between patients and controls increases the sample size requirements. The factors that are important in determining

### Table 1. Characteristics of Patient and Control Groups

<table>
<thead>
<tr>
<th>Subject Group</th>
<th>Age, y</th>
<th>Interval Between Scans, d</th>
<th>MMSE Score at First Scan</th>
<th>Brain Volume at First Scan, mL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer disease</td>
<td>65.0 (6.4), 55-76</td>
<td>336 (62), 215-434</td>
<td>19.6 (4.1), 11-26</td>
<td>1069.7 (142.4)</td>
</tr>
<tr>
<td>Controls</td>
<td>65.0 (10.5), 52-84</td>
<td>326 (90), 180-476</td>
<td>29.2 (1.0), 28-30</td>
<td>1175.2 (132.1)</td>
</tr>
</tbody>
</table>

*Unless otherwise indicated, data are mean (SD), range.
†MMSE indicates Mini Mental State Examination; maximum score, 30.

### Table 2. Necessary Sample Size per Treatment Arm Using Group Mean Rates of Atrophy Determined Using Serial MRI Registration in AD

<table>
<thead>
<tr>
<th>Magnitude of Treatment Effect to be Detected, % Reduction in Rate of Atrophy</th>
<th>Based on Rate of Atrophy in AD Group Alone</th>
<th>Controlling for Normal Aging (AD-Control Rates of Atrophy)</th>
<th>Also Assuming a 10% Patient Dropout</th>
<th>Also Assuming 10% of Scan Pairs Are Inadequate for Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>461</td>
<td>673</td>
<td>747</td>
<td>830</td>
</tr>
<tr>
<td>20</td>
<td>115</td>
<td>168</td>
<td>187</td>
<td>207</td>
</tr>
<tr>
<td>30</td>
<td>51</td>
<td>75</td>
<td>83</td>
<td>92</td>
</tr>
<tr>
<td>40</td>
<td>29</td>
<td>42</td>
<td>47</td>
<td>52</td>
</tr>
<tr>
<td>50</td>
<td>18</td>
<td>27</td>
<td>30</td>
<td>33</td>
</tr>
</tbody>
</table>

*MRI indicates magnetic resonance imaging; AD, Alzheimer disease. All data are number of subjects required.
sample size are first the difference between AD and control rates of loss for the structure to be measured; the maximum treatment effect that could be expected is to reduce the rate of loss to that of age-matched controls (100% disease slowing). The second factor is the heterogeneity or variability (the SD) of the measured rates of atrophy in the AD group. This variability is the result of genuine differences between individuals, dependent on stage of disease and so forth, as well as the measurement error of the technique used. The more homogeneous the measured rate of atrophy in the patients, the easier it is to detect a small treatment change.

Bearing these issues in mind, what then are the reasons for the wide range (6-fold) in the estimated sample sizes required by the different techniques? It may be that the patients included in some of the studies had intrinsically higher heterogeneity. The patients in our study were fairly carefully assessed reflecting a relatively pure group of clinically diagnosed probable AD. In addition, our patients as a group were younger than in many studies. While this may not affect the rate of atrophy in the AD group, it may mean that the controls, being younger, had lower rates of loss than might be seen in older controls. For instance, the mean age of the control group of Jack et al19 was 81 years. Rates of whole brain atrophy might similarly be higher in an older group. However, a recent article by Mueller et al25 suggests that healthy oldest-old subjects do not show greater rates of brain loss compared with younger elderly. These issues are controversial, and we are currently investigating how rates of cerebral atrophy in AD and in normal aging are affected by age.

Alternatively, techniques that measure brain substructures may be subject to greater measurement error because of the difficulty in manually segmenting (outlining) these structures. This may be particularly relevant for computed tomographic studies and for the earlier published studies where measurement techniques may have been less sophisticated. The use of manual methods of segmentation and volume measurement introduces the issue of interobserver and intraobserver variability. Measurement errors are likely to be greater for smaller and less easily defined neuroanatomical structures, particularly if an arbitrary cutoff to the structure is used (eg, the posterior boundary of the temporal lobe or hippocampus). Furthermore, a cerebral substructure may have a variable rate of loss as the disease progresses. For example, it may be that while the hippocampus is severely affected early in the disease, it may not lose tissue as rapidly in the progression from mild to moderate AD, whereas other areas may only show significant changes later in the disease. For instance, Jack et al19 found that in very mild AD (mean MMSE, 22/30), the mean hippocampal volume was already significantly smaller (1.75 control SDs) than that of controls. More severely affected patients (mean MMSE, 18/30) had a mean hippocampal volume that was 1.99 control SDs below the control mean—in effect only slightly smaller than that of the very mildly affected patients. By contrast, measuring total cerebral atrophy may even out these differences but lack sensitivity in very early cases.

The MRI registration technique addresses several of these issues. The measurement of total cerebral atrophy includes all areas of loss and avoids the necessity of making a priori decisions about relevant regions of interest. The technique has high repeatability: scan-rescan reproducibility based on subjects having 2 scans on the same day shows a mean absolute difference between the 2 measures of under 2 mL (<0.2% of brain volume). The technique is relatively insensitive to the operator because both registration of the scans and quantification of atrophy

<table>
<thead>
<tr>
<th>Source, y (Type of Study)</th>
<th>No. of Subjects Studied</th>
<th>Scan Interval, mo</th>
<th>Structured Measured</th>
<th>Mean (±SD) Change per Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AD</td>
<td>Controls</td>
<td>AD</td>
<td>Controls</td>
</tr>
<tr>
<td>Luxenber et al, 1987 CT</td>
<td>18</td>
<td>12</td>
<td>16</td>
<td>39</td>
</tr>
<tr>
<td>Burns et al, 1991 CT</td>
<td>63</td>
<td>0</td>
<td>12</td>
<td>NA</td>
</tr>
<tr>
<td>DeCarli et al, 1992 CT</td>
<td>20</td>
<td>17</td>
<td>13</td>
<td>20 F</td>
</tr>
<tr>
<td></td>
<td>36 M</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jobst et al, 1994 CT</td>
<td>61 (20†)</td>
<td>47</td>
<td>22 (17†)</td>
<td>19</td>
</tr>
<tr>
<td>Shear et al, 1995 CT</td>
<td>41</td>
<td>35</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AD indicates Alzheimer disease; CT, computed tomographic scan; MRI, magnetic resonance imaging scan; F, female; M, male; NA, not applicable; and CSF, cerebrospinal fluid.
†Subset with AD confirmed by histopathological analysis.
‡Initially no dementia observed to 12 months postdiagnosis.
are performed virtually automatically. The only operator input is in the preprocessing step, where scalp and nonbrain tissues are roughly removed in a semiautomated procedure that typically takes less than 10 minutes. The registration process also compensates for "scanner drift" or variation in voxel size; conventional volumetric MRI-based measurements do not allow for these variations.

Alzheimer disease is inevitably a heterogeneous condition, and rates of disease progression between subjects are very variable. An enhancement of this technique for monitoring the disease progression might be to use subjects as their own controls for rate of atrophy. This would require subjects to have a 3- to 6-month placebo run-in with 2 MRI scans to establish their individual rate of atrophy. A further MRI scan would then be performed after 6 to 12 months on either active treatment or placebo, with the outcome assessed by comparison with their baseline rate of atrophy.

In terms of practical application of the registration technique, our experience has shown that it is possible to successfully scan patients with MMSEs as low as 10/30. By collecting only the coronal volumetric sequence, it was possible to reduce the duration of the scan to as little as 10 minutes, which substantially improved tolerability. Techniques that reduce patient movement during scans (including faster scans) are very important in longitudinal studies because both of a subject's scans must be usable in order to calculate a rate of loss. The figure of 10% of unusable scan pairs is likely to be an underestimate unless great care is given to patient compliance and, where necessary, scans are repeated. It should be noted that a larger sample is needed if 10% of all scans (as opposed to scan pairs) were unusable and these scan failures were randomly distributed. For studies requiring 2 scans on each patient, this amounts to an increase of approximately 25%. For "within patient" trials requiring 3 scans, this amounts to an increase of approximately 35%. Improvements in MRI techniques mean that faster scanning with similar resolution is now possible, which, together with careful explanation of the scanning procedure, may reduce problems of movement artifact and patient dropout. These advances, together with improving image analysis, should mean that imaging studies of atrophy progression in AD are increasingly feasible in trials examining disease progression.

The absence of methods to measure disease progression at a molecular level has increased interest in using MRI-based measures of atrophy. However, there are some caveats to be attached to the use of this method as a surrogate marker of disease progression. First, cerebral atrophy is an indirect measure of pathological processes occurring at the cellular level. Second, therapeutic disease-slowing effects at the molecular or cellular level may in theory take some time to feed through to a slowing in the rate of atrophy. Nonetheless, macroscopic atrophy seems to be an inevitable concomitant of the cellular destruction, synaptic loss, and dendritic pruning of AD. If a therapeutic intervention in AD were to slow rates of cerebral atrophy as well as produce symptomatic benefit, this would be strong evidence for a disease-modifying effect. Registration of serial MRI offers a powerful method of tracking changes in rates of atrophy with potential for use in clinical trials.

### CONCLUSIONS

Serial MRI registration offers an accurate, powerful method for measuring rates of atrophy in AD. The results of this study suggest that the technique could be effectively used in a clinical trial setting evaluating the efficacy of a drug in modifying the biological progression of the disease. By comparison with the few published longitudinal imaging studies, quantification of total rates of cerebral atrophy based on registered MRI seems to require the smallest sample size, and may thus be the most cost-effective method available. Further research comparing longitudinal measurements of regional (eg, hippocampal) and global atrophy at different stages of the disease are needed to guide the choice of surrogate markers for forthcoming trials in AD.

Accepted for publication, April 26, 1999.

This study was supported by the Medical Research Council, London, England. (Dr Fox is supported by a Medical Research Council Clinician Scientist Fellowship. Dr Harvey is supported by an Alzheimer’s Disease Society Research Fellowship, London, England. Dr Rossor holds a Medical Research Council Program Grant.)

©2000 American Medical Association. All rights reserved.
REFERENCES


5. Corey BJ, Anand R, Veach J. A randomized trial evaluating the efficacy and safety of rasurgery, Queen Square, London WC1N 3BG (e-mail: research@dementia.ion.ucl.ac.uk).


24. Fox NC, Freeborough PA, Mekkaoui KF, Stevens JM, Rossor MN. Cerebral and cerebellar atrophy on serial magnetic resonance imaging in an initially symptom-free subject at risk of familial prion disease. BMJ. 1997;315:856-857.


