SHORT INTERVAL SERIAL MRI STUDIES IN ALZHEIMER’S DISEASE AND NORMAL AGEING

SUBMITTED TO THE UNIVERSITY OF LONDON

FOR THE DEGREE OF

DOCTOR OF MEDICINE

JONATHAN MARK SCHOTT

INSTITUTE OF NEUROLOGY, UNIVERSITY COLLEGE LONDON

2004
Abstract

This thesis investigates rate of brain atrophy derived from serial volumetric magnetic resonance (MR) scans as a measure of progression in Alzheimer’s disease (AD). Forty-six patients with AD and 23 controls attended for multiple (7-10) MR brain scans and two serial neuropsychological and clinical assessments over one year; some subjects had additional scans at 18 months and two years. Techniques for accurate scan comparison (registration) and automated atrophy quantification (using the brain boundary shift integral - BBSI) were assessed. Atrophy determined using the BBSI was compared with manual measures of brain and ventricular change. Significant differences in atrophy rate between patients and controls were shown at intervals as short as six months.

For each subject, the BBSI was used to calculate rate of whole brain atrophy from every possible scan pair; 2199 measurements were made for the patients, and 1182 for the controls. A multi-level model was used to determine mean atrophy rates (AD: 2.23%/year; controls: 0.49%/year) and inter- and intra-individual variances in rates for patients and controls. The improved precision of atrophy measurement made possible using the model was utilised to determine strategies for reducing sample sizes in clinical trials, and extended to determine factors both correlating with, and influencing, atrophy progression. Rate of cerebral atrophy was found to correlate with decline in a number of neuropsychological scores. Increasing disease severity, lower systolic blood pressure, and younger age at onset were found to predict subsequently increased rate of atrophy; the latter two factors appeared to be dependent on the former.

An unbiased, longitudinal group analysis was used to determine brain areas undergoing significantly increased regional atrophy in AD. The pattern of regional atrophy was found to alter with disease progression, and further evidence for a relationship between systolic hypotension, age of onset and disease severity was demonstrated.
1 Introduction ...................................................................................................................................17

1.1 Dementia ................................................................................................................................17

1.2 Alzheimer’s disease ................................................................................................................18

1.2.1 History and epidemiology .....................................................................................................18

1.2.2 Pathology .............................................................................................................................18

1.2.2.1 Histopathology .................................................................................................................18

1.2.2.2 Macroscopic changes ........................................................................................................20

1.2.3 Genetic determinants and risk factors ..................................................................................21

1.2.3.1 Familial Alzheimer’s disease ..............................................................................................21

1.2.3.2 Sporadic Alzheimer’s disease .........................................................................................22

1.2.3.2.1 Genetic risk factors ......................................................................................................22

1.2.3.2.2 Other risk factors .........................................................................................................23

1.2.4 Clinical Features ..................................................................................................................23

1.2.4.1 Atypical forms .................................................................................................................24

1.2.5 Diagnosis of AD ....................................................................................................................24

1.2.6 Quantitative assessments of functioning in AD ......................................................................25

1.2.6.1 Standardized mental test batteries ...................................................................................25

1.2.6.2 Cognitive Neuropsychology ............................................................................................26

1.2.7 Problems in achieving an accurate diagnosis ..........................................................................26

1.2.7.1 Detecting dementia ...........................................................................................................26

1.2.7.2 Excluding treatable causes of dementia .........................................................................27

1.2.7.3 Distinguishing AD from other causes of dementia ............................................................27

1.2.7.3.1 Dementia with Lewy Bodies .........................................................................................28

1.2.7.3.2 Frontotemporal Lobar Degeneration ............................................................................28

1.2.7.3.3 Vascular dementia .......................................................................................................29

1.2.8 Treatments for AD ................................................................................................................29

1.2.8.1 Future therapies ..............................................................................................................29
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2</td>
<td>Methods</td>
<td>80</td>
</tr>
<tr>
<td>6.3</td>
<td>Results</td>
<td>82</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Retention</td>
<td>82</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Clinical change</td>
<td>84</td>
</tr>
<tr>
<td>6.3.3</td>
<td>Neuropsychological change</td>
<td>86</td>
</tr>
<tr>
<td>6.4</td>
<td>Discussion</td>
<td>88</td>
</tr>
<tr>
<td>6.4.1</td>
<td>Retention</td>
<td>88</td>
</tr>
<tr>
<td>6.4.2</td>
<td>Clinical change</td>
<td>89</td>
</tr>
<tr>
<td>6.4.3</td>
<td>Neuropsychological change</td>
<td>91</td>
</tr>
<tr>
<td>7</td>
<td>Scan quality</td>
<td>92</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>92</td>
</tr>
<tr>
<td>7.2</td>
<td>Methods</td>
<td>92</td>
</tr>
<tr>
<td>7.3</td>
<td>Results</td>
<td>96</td>
</tr>
<tr>
<td>7.4</td>
<td>Discussion</td>
<td>97</td>
</tr>
<tr>
<td>8</td>
<td>A longitudinal case study of thalamic stroke</td>
<td>99</td>
</tr>
<tr>
<td>8.1</td>
<td>Introduction</td>
<td>99</td>
</tr>
<tr>
<td>8.2</td>
<td>Case report</td>
<td>99</td>
</tr>
<tr>
<td>8.2.1</td>
<td>Clinical assessments</td>
<td>99</td>
</tr>
<tr>
<td>8.2.2</td>
<td>Imaging</td>
<td>100</td>
</tr>
<tr>
<td>8.2.3</td>
<td>Neuropsychological assessments</td>
<td>101</td>
</tr>
<tr>
<td>8.3</td>
<td>Discussion</td>
<td>104</td>
</tr>
<tr>
<td>9</td>
<td>Stability in longitudinal scanning</td>
<td>106</td>
</tr>
<tr>
<td>9.1</td>
<td>Introduction</td>
<td>106</td>
</tr>
<tr>
<td>9.2</td>
<td>Methods</td>
<td>108</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Subjects</td>
<td>108</td>
</tr>
<tr>
<td>9.2.2</td>
<td>MRI acquisition, Phantom scanning and data analysis</td>
<td>108</td>
</tr>
<tr>
<td>9.2.3</td>
<td>Image processing</td>
<td>108</td>
</tr>
<tr>
<td>9.2.4</td>
<td>Recovering artificial scaling changes</td>
<td>109</td>
</tr>
<tr>
<td>9.2.5</td>
<td>Atrophy quantification in AD with and without scaling change</td>
<td>109</td>
</tr>
<tr>
<td>9.2.6</td>
<td>Comparison of TIV and registration using artificially scaled scans</td>
<td>110</td>
</tr>
<tr>
<td>9.2.7</td>
<td>Comparison of TIV and registration using patient and control scans</td>
<td>110</td>
</tr>
<tr>
<td>9.2.8</td>
<td>Statistical analysis</td>
<td>110</td>
</tr>
<tr>
<td>9.3</td>
<td>Results</td>
<td>110</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Recovering artificial scaling changes</td>
<td>110</td>
</tr>
</tbody>
</table>
Tables

Table 2-1. Annual longitudinal hippocampal atrophy rates in AD and controls........47
Table 3-1. Sample size estimates based on reported atrophy rates in AD ..............59
Table 4-1. Study design ..........................................................................................65
Table 5-1. Blood pressure of patients with AD and controls ...............................74
Table 5-2. Baseline cognitive scores .....................................................................76
Table 5-3. Expected and observed ApoE genotype vs mean age at onset ...........77
Table 6-1. Scanning schedule .............................................................................81
Table 6-2. Study visit attendance .......................................................................83
Table 6-3. Timing of achieved visits ....................................................................84
Table 6-4. Acetylcholinesterase inhibitor use ......................................................84
Table 6-5. Mean decline in MMSE score ..............................................................85
Table 6-6. Mean change in annual MMSE score and AchEI treatment ..........86
Table 6-7. Annualized mean (SD) change in neuropsychological score ..........87
Table 7-1. Qualitative assessment of scan quality ..............................................93
Table 7-2. Results of scan grading .....................................................................96
Table 7-3. Grading of scans performed on the same day in AD .......................97
Table 8-1. Baseline and one-year neuropsychology scores for ANC .............102
Table 10-1. Baseline ventricular volumes .........................................................122
Table 10-2. Annual change in ventricular volume (ml) over one year .............123
Table 10-3. Ventricular change over the two six month periods .................124
Table 10-4. Baseline ventricular-to-brain ratio ...............................................125
Table 10-5. Annual percent change in ventricular-to-brain ratio at one year ......125
Table 10-6. Ventricle-to-brain ratio annual percent change over six months ...126
Table 10-7. Reported absolute rates of ventricular expansion in AD (ml/yr) ....128
Table 10-8. Reported annual percentage increases in ventricular size in AD ....129
Table 11-1. BBSI derived atrophy vs. segmented volume differences .............139
Table 12-1 Ventricular change and BBSI-derived change over 6 and 12 months 142
Table 12-2. BBSI derived atrophy over multiple intervals from baseline (all scans) 143
Table 12-3. Atrophy rates calculated using the BBSI and “good” scans only ....144
Table 12-4. Sample size estimates for a variety of scan intervals from baseline 145
Table 12-5. Reported MRI brain atrophy rates and power calculations in AD ....148
Table 13-1. Atrophy over several different six month intervals .......................151
Table 13-2. Number of registrations performed. ..............................................................153
Table 14-1. Estimated sample sizes to power each arm of a one year study. ............171
Table 14-2. Estimated sample sizes to power each arm of a two year study ..........172
Table 14-3. Estimated sample sizes to power each arm of a six month study .......173
Table 15-1. Statistically significant longitudinal correlates of atrophy ...............177
Table 16-1. Statistically significant baseline predictors of atrophy ....................182
Table 17-1. Demographics of patients entering sub-group analysis.....................190
Figures

Figure 1-1. Schematic diagram to illustrate normal and abnormal APP processing...19
Figure 1-2. Postmortem brain from a patient with Alzheimer’s disease..................21
Figure 2-1. Registered coronal T1-weighted images of an individual with AD........41
Figure 2-2. Registration and hippocampal atrophy in mild AD..........................42
Figure 2-3. Outline of the steps required to quantify whole brain atrophy..............43
Figure 2-4. Voxel compression mapping in Alzheimer’s disease..........................44
Figure 3-1. Randomized withdrawal placebo-controlled design studies.................51
Figure 3-2. Randomized start placebo-controlled design studies........................52
Figure 3-3. Diagram to illustrate ideal properties of a surrogate marker.................53
Figure 3-4. Diagram to illustrate reasons surrogate markers may fail....................55
Figure 6-1. Graph to show longitudinal change in neuropsychological tests..........88
Figure 7-1. Examples of scan grading...................................................................94
Figure 7-2. Examples of scans with inadequate head coverage and wrap...............95
Figure 8-1. Demonstration of new thalamic lesion.............................................101
Figure 9-1. Artificial scaling factor plotted agains 9dof recovered scaling factor....111
Figure 9-2. Artificial scaling change correction using TIV and 9dof.......................112
Figure 9-3. 9dof registration volume scaling factor versus TIV ratio.....................114
Figure 9-4. Correction factors from the scanner QA and 9dof registration..........115
Figure 10-1. Baseline ventricular volumes for patients and control (ml)..............122
Figure 10-2. One year ventricular change for patients and controls....................123
Figure 11-1. The brain boundary shift integral....................................................134
Figure 11-2. t-statistics for the mean difference between BBSI and SVD.............137
Figure 11-3. z-statistics from a range of BBSI and SVD values............................138
Figure 13-1. Mean atrophy rate vs. mean inter-scan follow-up in AD..................154
Figure 13-2. Mean atrophy rate vs. mean inter-scan follow-up in controls.............154
Figure 13-3. Mean standard deviation against mean inter-scan follow-up in AD....155
Figure 13-4. Mean standard deviation vs. mean inter-scan follow-up in controls...155
Figure 13-5. Multiple direct and indirect measures of atrophy for one individual..157
Figure 13-6. Atrophy in an individual: the effect of poor scans..........................158
Figure 13-7. Observed mean atrophy rates vs. model-derived mean (line)..........164
Figure 13-8. Observed SDs atrophy rates vs. model-derived SDs (line)..............164
Figure 17–1. Patterns of regional atrophy in sporadic Alzheimer’s disease............. 193
Figure 17–2. Patterns of regional atrophy by baseline MMSE.................................. 194
Figure 17–3. Patterns of regional atrophy by baseline matrix reasoning z-score .... 195
Figure 17–4. Patterns of regional atrophy by baseline arithmetic z-score.............. 196
Figure 17–5. Patterns of regional atrophy by baseline composite cognitive z-score. 197
Figure 17–6. Patterns of regional atrophy by baseline systolic blood pressure..... 198
Figure 17–7. Patterns of regional atrophy by age at disease onset ..................... 199
Figure 17–8. Patterns of regional atrophy by disease duration............................ 200
Figure 17–9. Regional atrophy by MMSE corrected for matrices z-score .......... 201
Figure 17–10. Regional atrophy by systolic BP corrected for matrices z-score..... 202
Figure 17–11. Regional atrophy by age at AD onset corrected for matrices z-score. 203
Figure 17–12. Regional atrophy by AD duration corrected for matrices z-score..... 204

Appendix Figures
Figure A 1. Setting the upper threshold ................................................................. 222
Figure A 2. Setting the lower threshold ................................................................. 222
Figure A 3. Defining the inferior limit of the scan ............................................. 223
Figure A 4. Using erosions to define brain tissue .............................................. 223
Figure A 5. Using dilations to recover lost brain tissue .................................. 224
Figure A 6. Ventricular segmentation ................................................................. 225
Abbreviations

6dof Six degrees of freedom
9dof Nine degrees of freedom
AchEI Acetylcholinesterase inhibitor
AD Alzheimer’s disease
ADAS-Cog Alzheimer's disease assessment scale-cognitive subscale
ADRDA Alzheimer’s disease and related disorders association
AIDS Acquired immune deficiency syndrome
ApoE Apolipoprotein E
APP Amyloid precursor protein
BBSI Brain boundary shift integral
BMI Body mass index
CAST Cardiac arrhythmia suppression trial
CDR Clinical dementia rating
CERAD Consortium to establish a registry for Alzheimer's disease
CIBIC-Plus Clinician interview based impression of change plus caregiver input
CPALT Camden paired associate learning test
CSF Cerebrospinal fluid
CT Computed tomography
DLB Dementia with Lewy bodies
DSM-IV Diagnostic and statistical manual of mental disorders – 4th edition
fMRI Functional magnetic resonance imaging
FDA Food and drugs administration
FLAIR Fluid attenuation inversion recovery
FTD Frontotemporal dementia
FTLD Frontotemporal lobar degeneration
GRASS Gradient recalled acquisition at steady state
Hippo Hippocampus
HIV Human immunodeficiency virus
MCI Mild cognitive impairment
MIDAS Medical information display and analysis system
MMSE Mini-mental state examination
MR Magnetic resonance
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MTL</td>
<td>Medial temporal lobe</td>
</tr>
<tr>
<td>MRS</td>
<td>Magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>NART</td>
<td>National adult reading test</td>
</tr>
<tr>
<td>NIA</td>
<td>National institute for aging</td>
</tr>
<tr>
<td>NINCDS</td>
<td>National institute of neurological &amp; communicative disorders &amp; stroke</td>
</tr>
<tr>
<td>PA</td>
<td>Progressive (non-fluent) aphasia</td>
</tr>
<tr>
<td>PET</td>
<td>Positron emission tomography</td>
</tr>
<tr>
<td>PS1</td>
<td>Presenilin 1</td>
</tr>
<tr>
<td>PS2</td>
<td>Presenilin 2</td>
</tr>
<tr>
<td>QA</td>
<td>Quality assurance</td>
</tr>
<tr>
<td>RF</td>
<td>Radio frequency</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-noise ratio</td>
</tr>
<tr>
<td>SPECT</td>
<td>Single positron emission tomography</td>
</tr>
<tr>
<td>SPM</td>
<td>Statistical parametric mapping</td>
</tr>
<tr>
<td>SVD</td>
<td>Segmented volume difference</td>
</tr>
<tr>
<td>T</td>
<td>Tesla</td>
</tr>
<tr>
<td>TIV</td>
<td>Total intracranial volume</td>
</tr>
<tr>
<td>TE</td>
<td>Time to echo</td>
</tr>
<tr>
<td>TR</td>
<td>Time to repeat</td>
</tr>
<tr>
<td>VCM</td>
<td>Voxel compression map(ping)</td>
</tr>
<tr>
<td>VD</td>
<td>Vascular dementia</td>
</tr>
<tr>
<td>vMRI</td>
<td>Volumetric MRI</td>
</tr>
<tr>
<td>WASI</td>
<td>Wechsler abbreviated scale of intelligence</td>
</tr>
</tbody>
</table>
The Problem

Alzheimer’s disease (AD) is the commonest form of dementia in the developed world. AD inevitably follows a progressive course, leading to devastating loss of cognitive function and a reduced lifespan. As age is the single biggest risk factor for the development of AD, rising life expectancy in the western world means that AD will have increasingly profound implications not only for individuals and families, but also for society as a whole. The aetiology of sporadic AD is not yet known, but the disease is thought to result from a combination of environmental and genetic factors. Already the first drugs impacting on the symptoms of AD are widely available. However, as our understanding of the molecular basis of this disease improves, so does the possibility of rational, targeted drugs that may slow the pathological progression of the disease. Assessing whether such new compounds have disease-modifying properties represents a significant challenge. To date functional and neuropsychological scores have been used as outcome measures in clinical trials in AD. These tests are not without their problems; they have ceiling and floor effects, are non-linear, and may be open to subjectivity. Perhaps most importantly, scores on these tests may be highly variable both from day-to-day and examiner-to-examiner. These factors lead to substantial variances in test results meaning that large numbers of patients need to be studied in order to determine a symptomatic effect. Furthermore, using these measures alone, it is difficult to distinguish symptomatic from truly disease-modifying effects.

The pathological hallmarks of AD are excess accumulation of extra-cellular amyloid plaques, and intracellular neurofibrillary tangles. A truly disease modifying agent is likely to preserve neurons and is probably also likely to alter the density of these proteins, either by slowing their deposition, or by enhancing their clearance. It is not yet possible to assess the distribution or quantity of these abnormal proteins reliably in vivo. However an inevitable downstream effect of neuronal loss is cerebral atrophy. This atrophy is seen at post-mortem, but may also be visualized in vivo using volumetric magnetic resonance imaging (MRI). In AD, atrophy is both global, as well as regional, with early and prominent involvement of medial temporal lobe structures, including the hippocampus and entorhinal cortex. Cerebral atrophy is thus a good candidate for a surrogate marker of AD progression. In order to quantify rates of brain
atrophy at an individual level, it is necessary to compare serial scans taken from the same individual over time. In order to facilitate accurate comparison, these scans can be digitally matched (registered) to one another, with a high degree of precision. It is then possible to measure rates of global atrophy either by comparing the brain volumes directly, by assessing expansion of cerebrospinal (e.g. ventricular or sulcal) fluid space over time, or by using an automated method of brain volume change, the brain boundary shift integral (BBSI). Atrophy rates derived using the BBSI can distinguish groups of patients with Alzheimer’s disease from controls over a one year period. This difference in rate of global or regional brain atrophy between patients and age matched controls provides a useful signal that can be utilised as an outcome measure in a clinical trial. If a novel therapy were to have disease modifying properties, it would be expected that the rate of cerebral atrophy in the treated population would slow compared to those on placebo. This could provide evidence that the drug was having disease modifying, rather than purely symptomatic effects.

The aim of the work undertaken for this thesis was to determine the utility of whole brain atrophy rates derived from short interval scans (over periods of one year or less) and combinations of short interval scans as potential outcome measures for use in treatment trials of AD. Specific aims were: (1) to assess patient tolerability of multiple scans; (2) to assess and derive strategies to minimize sources of error related to image acquisition or analysis of serial scans; (3) to compare manual (ventricular) and automated (BBSI) measures of atrophy as progression markers over short intervals; (4) to provide a statistical framework to model BBSI derived short interval measures of atrophy, and thereby to determine within- and between-subject variability in rates of progression to assess whether changes in an individual’s rate of decline offers statistical advantages at trial end points; and (5) to use novel image analysis techniques to determine the patterns of progressive global brain atrophy occurring in sporadic AD. In this way this study aimed to determine the minimum intervals, number of scans and patients required to detect significant progression in AD over short intervals, and to provide an unbiased assessment of the regional changes occurring in sporadic AD over a one-year period.
1 Introduction

1.1 Dementia

Dementia is a clinical state defined as an acquired, usually progressive impairment of multiple domains of cognition including memory, in the presence of normal consciousness. The cognitive impairments must impact on normal social functioning, or the ability to continue in normal employment (Eastley and Wilcock, 2000). The most commonly used criteria for the diagnosis of dementia are those from the Diagnostic and Statistical Manual for Mental Disorders (DSM-IV) (American Psychiatric Association, 1994) (Appendix 1).

The estimated prevalence of dementia, whilst differing between studies, clearly increases with age. Meta-analyses of studies done in developed countries have established dementia prevalence at around 1.5% at age 65 years, doubling every 4 years to reach about 30% at 80 years (Ritchie and Lovestone, 2002). In the United Kingdom, the Government's Actuary Department estimated life expectancies in 2002 to be 76 years for men and 80 year for women (Government's Actuaries Department (UK), 2004); these are expected to increase in years to come. It has been estimated that by 2050, there will be 13.2 million patients with AD in the United States of America (Hebert et al, 2003). Dementia impacts not only on those affected, but on families, carers and society, which is increasingly looked upon to provide support and care. It is clear therefore that dementia will become an ever increasing social issue in years to come.

Dementia is a clinical syndrome and not a diagnosis, and there are many diseases that can cause, or be associated with dementia. In the elderly population, degenerative causes are the most prevalent, and, in the UK, Europe and North America, Alzheimer's disease (AD) is the common cause (Ritchie and Lovestone, 2002).
1.2 Alzheimer’s disease

1.2.1 History and epidemiology

Almost one hundred years ago Alois Alzheimer described plaques, neurofibrillary tangles, and arteriosclerotic changes in the neocortex of a woman with presenile dementia (Alzheimer, 1906). Kraepelin subsequently ensured that this presenile form of degenerative dementia should bear Alzheimer’s name (Kraepelin, 1910). This term continued to be restricted to severe forms of presenile dementia with abundant plaques and neurofibrillary tangles until the 1960s when it was determined that the clinical and neuropathological differences between the presenile and senile manifestations of primary degenerative dementias were insufficient to define separate diagnostic entities (Forstl, 2000). This finding prompted a change in view, such that senile dementia could no longer be accepted as an inevitable consequence of normal ageing, but was in fact a disease state.

1.2.2 Pathology

1.2.2.1 Histopathology

Whilst advances in histopathology have given many new insights into the pathology of Alzheimer’s disease, the amyloid plaque and neurofibrillary tangle first described by Alzheimer remain the histologically defining features of the disease. Modern histological techniques have now determined that the *neurofibrillary tangle* is composed of hyperphosphorylated tau (for review see Ritchie and Lovestone (2002)). Tau is a ubiquitous protein expressed in nerves, contributing to their structural and functional integrity. It exists in six different isoforms, three of which have three repeats in the extracellular domain (so called three-repeat tau), with the remaining three bearing four (four-repeat tau). Alzheimer’s disease is associated with hyperphosphorylation of both three- and four-repeat tau, distinguishing it at the molecular level from diseases associated with only three-repeat (e.g. Pick’s disease) or only four-repeat tau (e.g. corticobasal degeneration and progressive supranuclear palsy). The extracellular *amyloid plaque* is produced by cleavage of the amyloid precursor protein (APP). This protein can be cleaved by (at least) three different enzymes (Figure 1-1). Normally, cleavage occurs via the enzyme α-secretase, and to a lesser extent by β-secretase, both pathways resulting in the production of a non-amyloidogenic protein product. The abnormal, pathological process in AD is
associated with sequential cleavage of APP by $\beta$- and $\gamma$-secretase producing an abnormal protein product, 42 amino acids in length ($\alpha\beta$-42), the major constituents of the extracellular amyloid plaque (Mudher and Lovestone, 2002).

Figure 1-1. Schematic diagram to illustrate normal and abnormal APP processing.

The precise process by which amyloid and tau are overproduced, interact and lead to the development of AD is unclear. Some researchers suggest that the predominant process is the production of $\alpha\beta$-42: the so-called amyloid hypothesis (Figure 1-1) (Hardy and Higgins, 1992). This theory is supported both by the finding that mutations causing autosomal dominant AD all cause an increase in production of the $\alpha\beta$-42 fragment; in the case of transgenic mice this increase is often in the absence of hyperphosphorylated tau. This theory has led to the search for drugs that either block the $\beta$- and $\gamma$-secretase enzymes, or enhance $\alpha$-secretase, as a means of decreasing production of the aberrant proteins. The exact relationship between amyloid deposition and tau phosphorylation is unclear, although some authors have postulated mechanisms that might link the two (Mudher and Lovestone, 2002). Whatever the
cause of the abnormal protein deposition, a final common step is neuronal cell death. A consequence of cell death (and probably cell damage) is a reduction in the amount of neurotransmitter available. Together, these processes lead to the development and inevitable progression of cognitive decline typical of AD.

As both amyloid plaques and neurofibrillary tangles accumulate to a certain extent during normal ageing, histological criteria have been drawn up to aid the histopathological diagnosis of AD, and to distinguish the disease from normal ageing. These criteria include both the quantity of protein deposited, as well as the site and sequence in which they are deposited. Several criteria are in common usage. The Braak and Braak staging of AD is based on their work describing a postulated sequential progression of neuritic protein accumulation from transentorhinal cortex, to entorhinal cortex, hippocampus, and finally to the neocortex (Braak and Braak, 1991); the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria are based upon a semi-quantified age-corrected scale of plaque and tangle density (Mirra et al, 1991); and the NIA-Reagan criteria are based upon a probability statement of the topographic staging of plaques and tangles (Anon., 1997a).

1.2.2.2 Macroscopic changes

The accumulation of amyloid and tau within the brain is invariably associated with neuronal cell death, the result of which can be seen at post-mortem as brain atrophy. This progressive cerebral atrophy is an inevitable feature of AD. Atrophy in AD is global, affecting both cerebral hemispheres diffusely, resulting in cortical thinning, widening of sulcal spaces and increased size of the ventricular system (Figure 1-2). However, not all brain regions undergo atrophy at the same rate. The pattern of progression demonstrated by Braak and Braak is mirrored by the focality of the atrophy in AD, with medial temporal lobe structures, such as the entorhinal cortex and hippocampus being prominently affected in AD (Braak and Braak, 1991) (Figure 1-2).
Global brain atrophy is seen as enlargement of cerebrospinal fluid spaces and sulcal widening. The result of medial temporal lobe atrophy is shown (arrow).

Whilst other diseases are associated with different histopathology, excess neuronal death and atrophy are not only seen in AD but also in other neurodegenerative diseases. The focality of atrophy is often different in these diseases, as discussed later; thus, in the absence of histopathology, the pattern of regional atrophy, may help distinguish these conditions from one another (see chapter 2.2.4.2.2). Furthermore, a degree of cerebral atrophy is seen in normal ageing, as will be discussed later.

1.2.3 Genetic determinants and risk factors

1.2.3.1 Familial Alzheimer’s disease

Whilst the single greatest risk for AD is increasing age, a number of other factors are now known to be important in influencing an individual’s risk of developing the disease. A family history of late onset AD may increase the chances of developing AD approximately two-fold. In certain rare cases, the disease may be transmitted in autosomal dominant fashion. In these cases, the disease onset occurs at a younger age (often in the fifth or sixth decade), and causative mutations may lie in one of three genes detected to date: *presenilin 1 (PS1)* (on chromosome 14), *presenilin 2 (PS2)* (on
chromosome 1), and the *amyloid precursor protein* - *APP* - gene (on chromosome 21). Numerous mutations within these genes have been shown to be causative of AD; such mutations lead to the development of AD with ~100% penetrance, generally at younger ages than are typical for sporadic disease. Similarly, the majority (but not all) patients with Down's syndrome (trisomy 21) will develop features of Alzheimer's disease by their 40s, presumably as a consequence of carrying three copies of the APP gene (Jorm, 2000). Transgenic mice over-expressing one (or more) pathogenic mutation in the genes leading to autosomal dominant AD, are increasingly used as models of the disease, allowing an increased understanding of AD pathogenesis as well as providing a framework for the development of new diagnostic and therapeutic strategies (Hardy, 2004).

1.2.3.2 Sporadic Alzheimer’s disease

The epidemiology of the much more common, late onset, "sporadic" AD is complex; whilst age is the single biggest risk factor for the development of AD, it is likely that multiple combinations of risk factors, both genetic and environmental could lead to the development of the disease.

1.2.3.2.1 Genetic risk factors

There is evidence for the involvement of certain genes in the development of sporadic AD. The association between sporadic late onset AD and apolipoprotein E4 (ApoE) first described in 1993 (Saunders et al, 1993; Strittmatter et al, 1993) has since been confirmed in numerous studies. Apolipoprotein E (ApoE) exists in three allelic forms: E2, E3, and E4. Caucasians carrying the E4 allele are three (heterozygotes) to eight (homozygotes) times more likely to develop AD than individuals without an E4 allele. Furthermore, possession of an ApoE4 allele lowers the distribution of age of onset (Strittmatter and Roses, 1996). Nonetheless, between 40–70% of patients with late onset AD do not carry the E4 allele, and significant proportions of the normal population carry the E4 allele without any cognitive impairment into the ninth decade, thus limiting its value as a diagnostic test on an individual basis. Other sites on chromosomes 6, 9, 10, and 12 have been proposed to harbour potential risk genes for AD, although specific genes have not yet been conclusively identified.
1.2.3.2.2 Other risk factors

Several studies have reported that vascular risk factors are also risk factors for AD. Thus obesity, diabetes mellitus, hypercholesterolaemia, mid-life systemic hypertension (Skoog and Gustafson, 2003; Skoog and Gustafson, 2002), high plasma homocysteine, low folate, low vitamin B12 (Clarke et al, 1998) and lower levels of holotranscobalamin (Refsum and Smith, 2003) may be risk factors for sporadic AD; furthermore cerebrovascular disease may act to worsen cognitive performance in the earliest stages of Alzheimer's disease (Esiri et al, 1999). Other proposed risk factors, not all of which have been definitely proven, include: low education, low premorbid intelligence, head trauma, and low levels of testosterone in males (Hogervorst et al, 2002); by contrast in women higher levels of oestrogen may be protective (Hogervorst and Smith, 2002; Cutter et al, 2003; Norbury et al, 2003).

1.2.4 Clinical Features

Alzheimer’s disease typically starts in later life, with a progressive, insidious onset of memory impairment. This memory impairment is typically for episodic material, which may be considered as an individual’s “personal diary”. Thus commonly patients will lose track of day-to-day events, forget appointments and what they have done over the preceding days. The neural correlate for episodic memory are the medial temporal lobe structures, particularly the hippocampus and entorhinal cortex, structures known from pathological studies to be affected early in the disease process (Braak and Braak, 1991). At the earliest stages of the disease, patients may not fulfil criteria for dementia, having only memory impairment, without involvement of other cognitive domains or disruption of everyday activities. This intermediate stage has been referred to as amnestic mild cognitive impairment (MCI) (Petersen et al, 1999); although not all patients fulfilling criteria for MCI will develop either dementia or AD, approximately 12% of patients with MCI will convert to a diagnosis of AD per year.
As AD pathology progresses to involve other parts of the medial temporal lobe, and then to neocortical structures, the burden of cognitive impairment increases. Patients develop semantic and other linguistic impairments plus emerging difficulties with visuospatial and visuoperceptual tasks (Lambon Ralph et al, 2003), and often become withdrawn and lose confidence. As the disease progresses, patients become more and more cognitively impaired, eventually becoming entirely dependent. Later in the disease course seizures and myoclonus are not infrequently encountered.

1.2.4.1 Atypical forms

Whilst for the majority of patients with AD early memory impairment and subsequent loss of other cognitive domains is the norm, there is considerable variation in the pattern and progression of disease within patients. In some patients, this variation is such that they are said to have *atypical Alzheimer’s disease*. Several forms of atypical AD are recognized; these include patients presenting with predominant signs of biparietal dysfunction (apraxia, visuoperceptual and visuospatial impairment, dyscalculia and poor spelling) in the face of relative preservation of memory: *biparietal AD*. Other patients can present with prominent aphasia, or rarely progressive visual dysfunction (Galton et al, 2000). Such patients may present particular diagnostic challenges, as biparietal AD may resemble corticobasal degeneration, and patients with prominent aphasia may resemble the progressive non fluent subtype of frontotemporal lobar degeneration (see 1.2.7.3.2)

1.2.5 Diagnosis of AD

A definitive diagnosis of AD can only be made on a pathological basis, according to the criteria outlined in 1.2.2.1. Whilst on very rare occasions a brain biopsy is undertaken permitting a definitive histopathological diagnosis during life, for the most part, AD must be diagnosed clinically. Research criteria have been established by a number of groups, including the American Psychiatric Association (American Psychiatric Association, 1994), and the Work Group of the National Institute of Neurological and Communicative Disorders (NINCDS) the Alzheimer’s Disease and Related Disorders Association (ADRDA) (McKhann et al, 1984). The stated intention of the NINCDS work Group was to “establish and to describe clinical criteria for the
diagnosis of Alzheimer’s disease of particular importance for research protocols…”. These criteria, published in 1984, provide for three levels of diagnostic certainty: definite, probable and possible (Appendix 2). As the specificity for detecting AD is substantially lower with “possible” AD (Hogervorst et al, 2000), the studies that are described in this thesis were restricted to patients fulfilling NINCDS-ADRDA criteria for “probable AD”. In a small number of cases (3/3) the diagnosis has been confirmed as “definite AD” at post-mortem.

1.2.6 Quantitative assessments of functioning in AD

The clinical features of AD described above can often be elicited by a careful history and bedside clinical assessment. A more rigorous approach is to use standardized measures of cognitive function in order to quantify an individual’s level of functioning. This allows both for comparison between individuals, and with serial assessments, for quantification of change.

1.2.6.1 Standardized mental test batteries

A number of rating scales in AD have been developed for use in clinical or research practice. The mini-mental state examination (MMSE) (Folstein et al, 1975) was developed as a quick screening test for AD, and is in common clinical use (Appendix 3). The score ranges from 0 – 30, with broad cut-offs of >26: no impairment; >20 but <26: mild dementia; >12 but <20: moderate dementia; and <12 severe dementia. In a research setting, the clinical dementia rating (CDR) is used to assesses individuals in a range of cognitive and functional domains (Morris, 1993); a CDR score of 0 represents no impairment; 0.5 represents questionable dementia; 1 mild dementia; 2 moderate dementia; and 3 severe dementia (Appendix 4). Other scores used in the research setting include the Alzheimer's disease assessment scale-cognitive (ADAS-Cog) (Rosen et al, 1984), a battery that assesses a spectrum of cognitive functions commonly impaired in AD, with each such function being allotted a maximum score; higher scores indicate more severe impairment. The total score for this battery ranges from 0 (no impairment) to 70 (severe impairment). Patients with AD decline on average 7 to 9 points on this scale every year, although the rate of this decline varies considerably. A number of other cognitive and functional batteries are used to assess AD (for details, see (Hodges, 1994)).
1.2.6.2 Cognitive Neuropsychology

A more formal approach to the quantification of cognitive deficits is to use standard neuropsychological tests. Neuropsychological testing is based upon the fact that different brain regions subserve different functions, and that damage to specific brain areas can be detected and quantified using specifically designed tests (Hodges, 1994; McCarthy and Warrington, 1990). Such tests have been validated in the population to produce normal ranges, and thus assessments using a well selected battery of tests can determine the extent and pattern of an individual's degree of cognitive impairment compared to that expected for age; by comparing current performances with reading ability (which gives an estimate of premorbid ability relatively unaffected by the early stages of the disease), it may also be possible to determine whether there has been decline from an individual's expected level of functioning (Nelson, 1982). Serial assessment may be used to determine change over time, with the caveat that "practice effects" may need to be taken into account, particularly in high functioning individuals. Although neither 100% sensitive nor specific for a given disease, the pattern of cognitive impairment demonstrated using neuropsychology may give useful clues to the underlying disease process.

1.2.7 Problems in achieving an accurate diagnosis

In the absence of a definitive non-invasive diagnostic test, AD must be diagnosed on a clinical basis using standardized criteria (as above), backed up by appropriate investigations, including neuropsychological testing and neuroimaging. However the sensitivity and specificity of such approaches are not perfect and a degree of misdiagnosis is inevitable; even in the best centres this may be as high as 15% (Delacourte, 1998).

1.2.7.1 Detecting dementia

Standard criteria for AD state that patients must be impaired in at least one cognitive domain apart from memory, and that there must be impairment of activities of daily living (American Psychiatric Association, 1994; McKhann et al, 1984). It is therefore clear that patients with very early disease will not fulfil these criteria, and therefore cannot be diagnosed with AD. The concept of MCI (see above) has recently been
introduced in an attempt to recognize and define this stage of the disease. Formal neuropsychological testing may be invaluable in this setting, allowing for quantification of apparent and covert deficits, and allowing for progression of impairment (the hallmark of a degenerative dementia) to be demonstrated using serial assessments.

1.2.7.2 Excluding treatable causes of dementia

Potentially treatable or reversible causes of dementia include vitamin B12 or folate deficiency, structural brain lesions (including tumours and subdural haematomata), uraemia, hypothyroidism and depression, which all must be excluded before a diagnosis of AD can be made. Routine investigations to exclude these possibilities are therefore mandatory in all patients with cognitive impairment, and, according to the practice parameters of the Quality Standards Subcommittee of the American Academy of Neurology, should include: structural neuroimaging with either a noncontrast CT or MR scan, blood count, urea and electrolytes, glucose, folate, B₁₂ and thyroid function (Knopman et al, 2001). Whilst syphilis is a potentially treatable dementia, due to its rarity (at least in the United States), these guidelines do not recommend routine testing without specific reason. Where appropriate, electroencephalography may be useful to exclude temporal lobe seizure activity which can present with an amnestic syndrome mimicking AD (Hogh et al, 2002).

1.2.7.3 Distinguishing AD from other causes of dementia

Other degenerative dementias may mimic AD, and on occasion, in the absence of definitive in vivo diagnostic tests for any of these conditions, it may be difficult to distinguish between them. In vivo diagnosis therefore relies on attempts to distinguish the clinical phenotypes, often in combination with ancillary investigations. Diagnostic criteria have been drawn up for many of these conditions, in an attempt to aid their accurate definition. Such criteria are however neither 100% sensitive nor specific for any given diagnosis. The three commonest conditions which may be mistaken for AD are dementia with Lewy bodies (DLB), frontotemporal lobar degeneration (FTLD), and vascular dementia (VD)
1.2.7.3.1 Dementia with Lewy Bodies

Dementia with Lewy bodies (DLB) may account for 20% of dementia seen in elderly patients. The pathological features are deposition of Lewy bodies (composed principally of alpha-synuclein) within the neurones of the substantia nigra and cortex, leading to cell loss. Clinical features may be similar to AD with progressive cognitive and memory impairment, but also include prominent day-to-day fluctuations, visual hallucinations and parkinsonism (For review see McKeith et al (2004)). Consensus criteria to aid the diagnosis of DLB were published in 1996 (McKeith et al, 1996).

1.2.7.3.2 Frontotemporal Lobar Degeneration

Frontotemporal lobar degeneration (FTLD) comprises three prototypic syndromes: semantic dementia (SD), frontotemporal dementia (FTD), and progressive non-fluent aphasia (PA). The pathology in FTLD is varied, and comprises true Pick’s disease (3-repeat tau Pick bodies at histology); tau mutations (frontotemporal dementia with parkinsonism linked to chromosome 17); neuronal loss and microvacuolation; and ubiquitin-positive, tau-negative inclusion body histology. There is no consistent relationship between histopathology and clinical features, which appear to be best related to the site of focal atrophy. Thus FTD is prominently a behavioural syndrome, leading to either disinhibition or apathy, and is associated with frontal lobe atrophy; SD manifests as a disorder of word meaning, and is associated with asymmetric (dominant) temporal lobe atrophy; and PA is a disorder of expressive language associated with dominant hemisphere posterior frontal/anterior temporal lobe atrophy (for a review of FTLD, see (Tolnay and Probst, 2001). There is however considerable overlap between the different subtypes in terms of clinical features and distribution of atrophy. Consensus diagnostic criteria for FTLD were published in 1998 (Neary et al, 1998); in clinical practice distinguishing the various forms of FTLD from one another, and FTLD from AD, rests on clinical assessment, backed up by neuropsychological testing and imaging, which may demonstrate focal atrophy in FTLD.
1.2.7.3.3 Vascular dementia

Vascular dementia (VD) is a common cause of cognitive impairment, often occurring in combination with other forms of dementia – so called mixed dementia. Vascular dementia may occur secondary to cortical infarction; multiple small subcortical infarcts; diffuse white matter disease; or distinct small infarcts disrupting cortico-subcortical connections (e.g. within the thalamus). The presence of vascular risk factors including smoking, hypertension, hyperlipidaemia or family history, and neurological signs consistent with upper motor neuron involvement may suggest a diagnosis of VD. Investigations including neuropsychology and imaging revealing ischaemic lesions (white matter hyperintensities on T2 or FLAIR MR imaging) may aid the diagnosis (for review, see (Roman, 2003)). For research purposes, consensus criteria for vascular dementia were published in 1993 (Roman et al, 1993).

1.2.8 Treatments for AD

Over the last decade, the first symptomatic treatments have been made available for AD. The acetylcholinesterase inhibitors (AchEI) exert their action by inhibiting the enzyme acetylcholinesterase, thus increasing brain availability of the neurotransmitter, acetylcholine. Several studies have demonstrated their efficacy in mild-to-moderately advanced AD, and currently three drugs are licensed in the UK: Donepezil, Rivastigmine and Galantamine (Doody, 2003). Within the last year, Memantine, a centrally acting NMDA antagonist, has been the first drug advanced for advanced AD in the UK (Rogawski and Wenk, 2003). The availability of these drugs has increased the need for early and accurate diagnosis of AD; thus, whilst AchEI often improve symptoms in DLB (McKeith et al, 2004) and mixed (VD and AD) dementia, there is anecdotal evidence to suggest that such drugs may worsen some of the symptoms associated with FTLD.

1.2.8.1 Future therapies

To date all drugs licensed for the treatment of AD are symptomatic therapies; no drugs have yet been shown to slow the progression of the disease process. However, with an increased understanding of the molecular pathology and genetics underlying AD, disease modifying therapies are now an increasingly realistic prospect. Recent
studies demonstrating clearance of amyloid plaques from the brains of transgenic AD mice (Schenk et al, 1999) led to a phase three study of amyloid vaccination in humans, which was halted due to the development of encephalitis in a number of patients (Munch and Robinson, 2002). Nonetheless, altered vaccination strategies may yet prove useful, as may drugs designed to block β and γ-secretase, enhance α-secretase or promote systemic amyloid removal, all of which are currently in development. A major challenge is to determine the efficacy of such strategies in clinical trials. Structural neuroimaging using MRI has been proposed as a means of assessing the disease-modifying properties of new drugs in AD, and this was the motivation for the work in this thesis.
2 Neuroimaging in Alzheimer's disease

Neuroimaging relevant to AD can broadly be divided into structural and functional techniques. Structural imaging comprises computed tomography (CT) and magnetic resonance imaging (MRI), and can provide a highly detailed image of the brain and cerebral structures, and permit differentiation of tissue types. A related modality, magnetic resonance spectroscopy (MRS) can provide a measure of a number of metabolite concentrations, which may alter as a consequence of AD. Functional techniques include positron emission tomography (PET), single photon emission computed tomography (SPECT) and functional MRI (fMRI) which all provide measures of cerebral metabolism and blood flow. This thesis provides data based on serial imaging using structural MRI; other imaging modalities (e.g. CT) will only be considered briefly with reference to previously reported studies.

2.1 Magnetic resonance imaging

Over the past two decades, MRI has established itself as a safe, non-invasive and high resolution means of imaging the brain which is increasingly available in clinical practice. Even a decade ago MRI was described as the "gold" standard for brain imaging (Scheltens et al, 1992); since that time technological advances have improved the resolution of MRI and the speed with which scans may be obtained, ensuring that MRI is still considered the pre-eminent structural neuroimaging modality (Scheltens et al, 2002).

2.1.1 MRI acquisition

MR imaging utilises the fact that protons in a magnetic field emit a radio signal following excitation by a radiofrequency (RF) pulse. In its simplest terms, MR can be considered as follows: (1) a patient is placed within the magnetic field in a scanner; (2) a RF pulse is delivered which excites the protons; (3) the RF pulse is turned off and the excited protons return to their resting state, releasing energy as they do so; (4) this energy signal is received and used to reconstruct the image. Since every cell of the body contains water (and thus protons), an MR image theoretically can provide information relating to every cell in the volume scanned. In practical terms, routine
MR scans generate magnetic fields ($B_0$) which vary from 0.3 – 1.5 Tesla (T), although higher field strength units are available. Unpaired protons become aligned with this field depending upon the strength of the magnetic field and their thermal energy. Protons within a magnetic field precess (spin) in the direction of the field with a frequency ($f_0$) determined by the Larmor equation:

$$f_0 = gB_0$$

where $g$ is the constant defined by the magnetic property of the nuclei. When a short RF pulse (of the order of 50mT) is transmitted at the resonance frequency of the proton, these protons absorb energy and move out of their original alignment both to a higher energy level (decreasing longitudinal magnetisation) and such that they precess in phase (establishing transverse magnetisation). Once the RF pulse is switched off, the protons subsequently realign with the magnetic field and emit energy in the form of radio-waves. The subsequent increase in longitudinal magnetisation gives rise to the so called T1-curve (the time constant T1 is the longitudinal or “spin lattice” relaxation time); the loss of transverse magnetisation gives rise to the T2-curve (the time constant T2 is the transverse or “spin-spin” relaxation time and is much shorter than the T1 time constant). The radio waves thus emitted can be converted into spatial information as the magnetic field within the scanner is not uniform, but instead is generated as a gradient, ensuring that emission of radio waves is additionally dependent upon the position of protons within this gradient field.

Different tissues have different MR properties dependent upon a number of factors including the number of free protons and how tightly bound they are. By altering acquisition parameters, the different T1 and T2 properties of different tissues can be utilized to generate an image highlighting a particular tissue type. The time to repetition (TR) is the time period between RF pulses; the longer the TR, the more likely it is that all protons will have full recovered to their baseline state. A shorter TR allows the differential T1 properties of different tissues to be utilized in generating an image, as different tissues will be in different states of excitation. It follows that short TR times improve T1 tissue contrast. Time to echo (TE) is the time between transmission of the RF pulse and collection of the signal; a short TE thus gives more T1 and less T2 contrast. In practical terms, a short repetition time leads to T1
weighting (in which the CSF appears dark); a long TE leads to T2 weighting (in which CSF appears white).

In addition, alterations of TR and TE will have other effects on the scan and scan quality. Signal-to-noise ratio (SNR) will be increased with longer TRs, shorter TEs, increased voxel size, and increased field strength. Scan time will be increased with longer TR, smaller voxel size and larger field of view. Resolution is increased with smaller voxel sizes, with the resulting decrease in SNR and increase in scan time. There is inevitably a trade-off between these various parameters, dependent upon the type of imaging required. Furthermore, data acquisition can be performed in different imaging planes and in three dimensions (volumetric MRI). Volumetric scans are T1 weighted and generally consist of around one million three dimensional voxels (each ~1mm³) in an adult brain, each with a different contrast, providing high anatomical resolution. Such scans are ideally suited to quantification of cerebral atrophy; volumetric MRI is therefore the imaging modality that is the focus of this thesis.

MRI is well established as a safe means of acquiring repeated images without the ionizing radiation of CT. In view of the strong magnetic field, MRI is however contraindicated in patients with metallic implants (e.g. cardiac pacemakers and intracerebral clips). Some subjects find MRI too claustrophobic to tolerate, and care must be taken to provide adequate ear protection, as the noise generated by the scanner gradients may be associated with cochlear damage (Radomskij et al, 2002).

2.1.2 MRI scan quality

The quality of MRI depends not only upon the scanning parameters chosen, but also on a number of other factors. The commonest cause of subject related artefact is movement, which appears as blurring of the image. This may be due to patient movement in the scanner, or due to involuntary physiological movement e.g. pulsatile blood flow. MRI acquisition related artefacts include inadequate head coverage, when brain tissue is not completely included within the field of view. Related to this is wrap (or aliasing) which occurs when part of the anatomy extends outside the field of view, but is still inside the area of radiofrequency excitation. This commonly results in the edge of the left hand side of the image being overlaid on the right; i.e.
signal from the subject’s nose being overlaid on the back of the brain. Finally, a major artefact caused by imaging hardware is scan inhomogeneity, seen as a slowly varying intensity gradient caused by spatial variation in the image signal. Developing methods to reduce and counter these problems is an essential part of any clinical trial utilizing MRI as an outcome measure.

2.2 Cross-sectional structural imaging in AD

2.2.1 Excluding treatable causes

The traditional role of imaging in AD (or other dementias) has been to exclude potentially treatable causes of dementia, such as tumours, hydrocephalus and subdural haematoma. Whilst detection of such a lesion is rare (Alexander et al, 1995; Scheltens et al, 2002), the possibility of a treatable cause must clearly always be excluded before the diagnosis of a degenerative dementia can be considered. Both CT and MRI are generally sufficient to demonstrate such lesions (Knopman et al, 2001), and in this context, the choice of imaging modality generally depends upon availability.

2.2.2 Diagnosis

More recently, structural imaging, and particularly MRI has been used not only to exclude treatable causes of dementia, but also to begin to provide a means of differentiating AD from either normal ageing or other degenerative dementias – imaging “beyond exclusion” (Scheltens et al, 2002). Neuroimaging is increasingly being used to add support to a clinical diagnosis in AD, and, in vascular dementia, imaging already forms a mandatory part of some clinical criteria. Whilst current guidelines do not recommend the use of imaging in the positive diagnosis of dementia (Knopman et al, 2001), in the correct setting, neuroimaging may be useful in this regard.
2.2.3 Normal ageing

If cross-sectional imaging is to be used to differentiate AD from normal ageing, then the brain changes and cross-sectional imaging features of AD must be compared with those associated with normal ageing. Early evidence about the effect of ageing on brain structure came from autopsy studies from the 19th century. This work suggested that brain weight declined gradually but only slowly (about 0.1% a year) from early adulthood until around the age of 60 (Blinkov and Glezer, 1968). Such studies were limited by the inclusion of few elderly people, a lack of clinical data, and were also confounded by the profound secular changes that were taking place in the 19th and early 20th centuries: as nutrition improved, populations had been growing taller and heavier. In the 1960s Miller and Corsellis (Miller and Corsellis, 1977) showed that the brain was not immune to these changes: brain weights increased by about 1 g (0.05%) a year for those born between the mid-19th and mid-20th centuries in the United Kingdom. Accounting for these changes led to the conclusion that brain weights were stable between the ages of 20 and 50 but fell progressively thereafter (Miller et al, 1980). Later studies with thousands of autopsies suggested that brain weight peaks by the mid-to-late teens and declines slowly (0.1–0.2% a year) until the seventh decade after which losses accelerate (Dekaban, 1978; Ho et al, 1980).

Early imaging studies using both CT and MRI implied that brain volume decreased and cerebrospinal fluid volumes increased with advancing age (Coffey et al, 1992; Jernigan et al, 1990; Pfefferbaum et al, 1994). Latterly, imaging studies using a variety of automated techniques designed to detect consistent group changes (computational neuroanatomy (Ashburner et al, 2003)) have been used to demonstrate non-linear changes in brain volume in normal ageing (Sowell et al, 2003), with preferential atrophy in prefrontal cortex (Raz et al, 1997; Thompson et al, 2001), and possibly relative sparing of medial temporal lobe structures (Good et al, 2001).
2.2.4 Alzheimer’s disease

As discussed in chapter 1.2.2.1, the histopathological progression of AD from medial temporal lobe structures through to involvement of more lateral medial temporal lobe structures and then to the neocortex has the macroscopic correlate of excess atrophy in these regions. The result of this excess atrophy may be visualized both using CT and (with a very high degree of resolution) using MRI. A number of studies have therefore assessed both global cerebral and medial temporal lobe atrophy as potential diagnostic markers of AD. In order for cross-sectional comparisons between individuals to be performed, it is necessary to correct such measurements for head size using an atrophy-invariate measure such as total intracranial volume (TIV) (Whitwell et al, 2001).

2.2.4.1 Computed tomography

A number of studies prior to 1990 assessed the role of CT measures of global brain atrophy as a diagnostic measure to distinguish AD from normal ageing; a review by DeCarli et al dating from this time showed that if two dimensional or three dimensional measurements were made, patients and controls could be distinguished with relatively high (63% to 88%) sensitivity for a given specificity (90%). Later studies suggested that measures of CSF volume, Sylvian fissure and temporal horn improved discrimination (Sandor et al, 1992; Sullivan et al, 1993).

Due to the relatively low resolution of CT and difficulties associated with the plane of acquisition, making accurate measurements of medial temporal lobe structures may not be straightforward using CT; nonetheless several CT studies demonstrated the utility of this methodology (Jobst et al, 1992a; de Leon et al, 1993; Jobst et al, 1992b), and were amongst the first to determine that in vivo structural measures could be useful aids to the accurate diagnosis of AD. A retrospective analysis of these and a number of subsequent CT studies assessing medial temporal lobe structures reported a high overall average sensitivity and specificity (84% and 92% respectively) in patients with AD of moderate severity (Frisoni, 2001).
2.2.4.2 MRI

The high resolution afforded by MRI has increasingly been harnessed as a potential diagnostic tool, both to distinguish AD from normal ageing, and from other neurodegenerative dementias. MRI based cross-sectional studies have assessed both measures of global brain atrophy (whole brain or ventricular volumes), and medial temporal lobe structures, as areas prominently affected in AD. Such techniques include visual assessment, linear measures, and the more anatomical precise but operator intensive volumetric measures, which involve painstaking outlining of a predetermined region of interest (ROI) on several sequential slices of a volumetric scan.

2.2.4.2.1 Distinguishing AD from normal ageing

Although the results of MRI based measures of ventricular volume comparing AD with normal ageing are broadly in keeping with those made using CT, there is evidence that MRI may be the superior modality (Sandor et al, 1992). Concomitant with increasing ventricular and sulcal spaces, total MRI brain volume measurements are reduced in AD compared with controls (DeCarli et al, 1996), and this loss correlates with the severity of cognitive impairment (Murphy et al, 1993).

A review of thirteen volumetric studies involving over 400 patients with mild-moderate AD, found that hippocampal and medial temporal structure volumes were able to detect dementia from normal ageing with an average sensitivity of 85% and specificity 88% (Bosscher and Scheltens, 2002). A similar analysis of 10 studies using visual ratings or linear measures revealed very similar results: an average sensitivity of 86% and specificity of 85% (Bosscher and Scheltens, 2002). These figures are similar to those found using CT measures of medial temporal lobe (MTL) atrophy (Jobst et al, 1992b), suggesting that in experimental settings, both simple and complex tools based on CT and MRI may be useful in distinguishing patients with AD from non-demented controls, with “reasonably high” accuracy (Frisoni, 2001).

Hippocampal volumetry using MRI has been used in an attempt to distinguish patients with MCI from normal ageing. Cross-sectional studies have suggested that measures of hippocampus or entorhinal cortex may be useful predictors as to which patients will
proceed to develop AD from either MCI or normal ageing (Du et al, 2001; Krasuski et al, 1998; Dickerson et al, 2001; De Santi et al, 2001; Rusinek et al, 2003). Furthermore studies of patients destined to get AD on the basis of a mutation in either the PS1, PS2 or APP genes have been shown to have reduction of medial temporal lobe volumes prior to the onset of symptoms (Fox et al, 1996b; Schott et al, 2003).

Whilst cross-sectional volumetric measures, particularly of medial temporal lobe structures appear to be useful tools in aiding the distinction of normal ageing from AD, it is important to note that these reports are based on research samples, and to date there is no validation of their use in routine clinical practice.

2.2.4.2.2 Distinguishing AD from other dementias

Cross-sectional MRI may also be useful in distinguishing AD from other forms of dementia. White matter disease on T2 weighted imaging may, in the correct clinical context, be suggestive of vascular or mixed dementia. Different patterns of atrophy (asymmetric and focal lobar) are commonly seen in FTLD (Chan et al, 2001) and may aid diagnosis in the appropriate setting. Certain rare dementias may be associated with imaging features which may help in their diagnosis, e.g. the pulvinar sign in variant Creutzfeldt-Jakob disease (Collie et al, 2003). However, in many conditions there may be no specific imaging features on routine CT or MRI to distinguish one dementia from another (e.g. DLB from AD), and in others neuroanatomical changes may be too mild, diffuse, or topologically complex to be detected by simple visual inspection or manually traced measurements of regions of interest. Powerful computerised methods are currently being developed in an attempt to aid differentiation of dementias on a cross-sectional basis, but to date, none is near acceptance in routine practice (Ashburner et al, 2003).

2.3 Longitudinal structural imaging in AD

Cross-sectional studies are by their nature limited: there is large inter-individual variation in brain size and structure; the progression and rate of atrophy can only be inferred from a one-off scan; and it must be assumed that progression of brain ageing is similar between individuals. Longitudinal studies, by using subjects as their own control, allow progression of atrophy to be quantified at the individual level. This in
turn provides the potential for longitudinal atrophy rates to be used as diagnostic tests, or as outcome measures for trials, under the premise that a truly disease-modifying drug would reduce atrophy rates towards that of normal ageing.

For these aims to be realistic: (1) it must be possible to compare serially acquired MRI scans from the same individual accurately and to apply robust techniques to measure volume change; (2) an accurate understanding of the changes associated with normal ageing is required; (3) and decisions may need to be made as to whether whole brain (global) or specific regions (e.g. hippocampus) should be assessed, and which technique(s) should be used to measure them.

2.3.1 Technical considerations

The accurate assessment of inter-individual differences or within-subject change over time, is aided by matching scans positionally so that they are in a common spatial framework. This needs to be performed post-acquisition, as with serial scanning subjects may not have been placed in exactly the same position within the scanner; this may be a particular problem in elderly patients with limited mobility. In addition there are inevitable changes in the MRI magnetic field that occur over time due to scanner gradient changes. These changes can result in fluctuation in voxel size, further complicating comparison (Whitwell et al, 2001; Gunter et al, 2003). Whilst every effort should be taken to ensure correct and consistent placement within the scanner, and to minimize hardware and software changes during longitudinal studies, another way to counter these problems is to correct the positioning differences and/or voxel drift by applying a post acquisition registration. The goal of registration is to align regions of interest on a baseline scan to the same region in follow-up scans, so that the voxel representing the same underlying anatomical structures are superimposed. This allows for direct comparisons of either global or regional brain structures. For the purposes of comparing structural scans and calculating rates of atrophy, rigid body registration, in which the baseline and repeat scans are considered to be rigid structures that must be matched onto one another without structural change, is the most commonly used technique.
2.3.1.1 Rigid body registration

A number of techniques have been designed to perform rigid body registration (For review see Ashburner et al (2003)). The most basic method is to identify a number of corresponding anatomical points on two consecutive scans and to determine the transformations that align these points accurately on top of one another. A minimum of three linearly independent points are required to determine the unique three dimensional coordinate system of each study allowing superimposition. The more points that are determined on the scan, the greater the accuracy of registration. However, manual definition of such points is time consuming and accuracy is limited by the ability of the user to define the anatomical structures reliably. Furthermore, the technique demands the existence of several invariate points; such lack of variation cannot be assumed in brains undergoing neurodegeneration.

Several automated techniques have been designed to allow accurate scan matching (Woods et al, 1998; Freeborough et al, 1996). These algorithms compare signal intensities from all voxels within the brain. By matching signal intensities from the voxels of the repeat to the baseline scan it is possible to improve the registration of intra-subject MR scans significantly. The aim is to superimpose voxels of a similar signal intensity (i.e. grey voxels to grey voxels; white voxels to white voxels), thereby using all the information within the scan to perform the matching process.

Freeborough and Fox described such a voxel intensity based method of registration, employing an optimisation procedure based on minimising the standard deviation of voxel intensity ratios (Freeborough et al, 1996). The basic registration involves: (1) segmenting (scalping) the scan to remove non-brain tissues (e.g. skull and scalp); (2) determining and applying a series of rotations and translations to match a repeat scan to the baseline. Three translations are applied in each direction within the x, y, and z axis and three rotations around these axes produce a so-called six degrees of freedom registration – 6dof. The process can be extended to allow voxels to stretch to a small extent in each of the three planes (nine degrees of freedom registration – 9dof). This theoretically allows for correction of inconsistencies in voxel size that may occur between the two scans. Once registered, a difference image can be created, so that change between scans can be visualized (Figure 2-1). This technique and its variants have been extensively validated (Freeborough et al, 1996; Paling et al, 2004; Gunter
et al, 2003), and are now widely used to facilitate accurate comparisons between serially acquired scans. However, the possibility that 9dof registrations may, by correcting for voxel drift, also "correct away" atrophy has not been critically assessed in studies to date.

Figure 2–1. Registered coronal T1-weighted images of an individual with AD. (a) Baseline image; (b) repeat scan registered to baseline; and (c) difference image (i.e. scan A subtracted from scan B) following rigid body registration.
2.3.1.2 Calculating atrophy

Once two serial scans have been registered to one another, outlining of the brain or brain substructures can then be performed on both baseline and repeat scans. When performed on sequential slices of a volumetric image, volumes of substructures (such as the hippocampus) can be accurately assessed, and atrophy rates calculated (Figure 2-2).

Figure 2-2. Registration and hippocampal atrophy in mild AD.
T1-weighted volumetric MRI scans were acquired at 18 month intervals; similarly positioned coronal slices are shown: the right panel is registered to the middle panel. The (right) medial temporal lobe is outlined (box), and the hippocampus outlined. Progressive global (upper panel) and hippocampal (lower panel) atrophy is shown. Note the high level of anatomic correspondence between the two registered scans (right and middle).

The same region-of-interest technique can be used to outline whole brain, or other specific brain substructures. However, any procedures involving outlining structures by hand are inevitably associated with measurement error; this is particularly the case for small structures such as the hippocampus and entorhinal cortex.
An automated technique, the brain boundary shift integral (BBSI) has been validated as a robust tool for quantifying global brain atrophy in degenerative dementia (Fox and Freeborough, 1997; Freeborough and Fox, 1997). Details of the algorithm are discussed in chapter 11; in brief, the BBSI calculates the shift at the brain/CSF boundary at every point across the three dimensional registered scan-pair, the sum of which approximates closely to the brain volume lost; being a direct measure this leads to a reduction in the error associated with indirectly comparing two scans (Freeborough and Fox, 1997). An overview of the steps required to calculate whole brain atrophy using the BBSI is shown in Figure 2-3 (after Paling et al, 2004).

Figure 2–3. Outline of the steps required to quantify whole brain atrophy.

After acquisition of two serial volumetric MRI scans, both must be segmented (scalped) - stage 1. The two segmented scans are then digitally matched using a rigid-body registration - stage 2. Quantification is performed using the brain boundary shift integral which sums voxel shift between scans 2 and 1 at every brain:CSF boundary - stage 3.
2.3.1.3 Non-linear registration

Non-linear registration allows for a more accurate matching of gyral anatomy by allowing scans not merely to match through movement as a rigid body, but additionally through a warping procedure. One such approach is fluid registration (Christensen et al, 1996; Freeborough and Fox, 1998), which models the transformation from one scan to another based on the physical properties of a viscous compressible fluid. Following a 9dof rigid body registration the model aims to find an exact match between source and target scan; in so doing a displacement vector is generated for each voxel within the image, with the degree of stretch or contraction demonstrating atrophy or expansion over the period; this can be demonstrated as an overlay image (voxel compression map - VCM), providing an unbiased means of demonstrating the pattern of regional atrophy occurring between two scans without the need for a priori decisions as to which structures should be assessed (Fig 2-4).

Figure 2–4. Voxel compression mapping in Alzheimer’s disease.
Regional atrophy (blue – green) is seen in the medial temporal lobes and parietal cortices. Expansion (red – yellow) is seen in CSF spaces including the sulci and ventricles.
This technique allows an individual's pattern of progressive atrophy to be visualized. An extension of this technique is to compare fluid-registered scans from individuals within a statistical framework (statistical parametric mapping – SPM (Friston et al, 1995)), allowing consistent patterns of increased atrophy within groups of patients to be demonstrated (Scahill et al, 2002). Further use of this technique is described in chapter 18. Fluid registration has also been shown in pilot studies to be useful in propagating outlined hippocampal volumes forward from a baseline scan to a registered repeat, thus potentially increasing accuracy and reducing time in making hippocampal measurements (Crum et al, 2001); however at present fluid registration is generally used to demonstrate qualitative rather than quantitative patterns of atrophy. Other computational methods for determining atrophy patterns in an unbiased fashion have also been developed (Sowell et al, 2003; Thompson et al, 2003); these and others are summarized by Ashburner et al (Ashburner et al, 2003).

2.3.2 Longitudinal volumetric imaging: results in normal ageing

Several longitudinal imaging studies have assessed the changes associated with normal ageing, typically over 1–5 years. These studies, with high-resolution MRI, show that rates of global atrophy in healthy people increase gradually with age from an annual rate of 0.2% a year at age 30–50 to 0.3–0.5% a year by the eighth decade (Resnick et al, 2003; Mueller et al, 1998; Scahill et al, 2003; Schott et al, 2003; Fox et al, 1999a). Longitudinal studies of apparently healthy individuals suggest that hippocampal atrophy rates increase from around 0.1–0.2% a year in those aged 30–50, to 0.8% in those in their mid-70s, rising further to 1.5–2% a year at 80–90 (Jack et al, 1998; Kaye et al, 1997; Scahill et al, 2003; Schott et al, 2003). This slow but definite atrophy with slow acceleration over time needs to be taken into account when assessing rates of atrophy in pathological states.

2.3.3 Longitudinal volumetric MRI studies in AD

Over the past few years, an increasing number of studies have reported atrophy rates in AD based on longitudinal studies. In these studies, both rates of whole brain atrophy, and atrophy of a wide range of structures have been compared using several different techniques. The majority have been performed using volumetric MRI which affords excellent resolution and thus anatomical specificity.
2.3.3.1 Global brain atrophy

Global brain volume loss can be assessed using manual measures of whole brain or CSF spaces or automated techniques (e.g. the BBSI). Longitudinal studies in AD invariably show better discrimination between patients and controls than cross-sectional analyses, which are confounded by the wide inter-individual variation in brain volume. Ventricular expansion is an inevitable consequence of atrophy of brain tissue, and can be reliably and accurately measured from serial scans. Studies suggest that ventricular expansion occurs at approximately 5-8ml/yr (DeCarli et al, 1992; Shear et al, 1995; Wang et al, 2002; Silbert et al, 2003), far higher than the 0.7-1ml per year seen in healthy controls (Schott et al, 2003; Scahill et al, 2003; Wang et al, 2002). Measures assessing whole brain change using several methods including the direct automated BBSI technique have determined rates of whole brain atrophy in AD to be approximately 2-2.4%/yr (Fox et al, 1999a; Fox et al, 2000; O'Brien et al, 2001; Bradley et al, 2002; Wang et al, 2002). Again these results are far in excess of those reported in normal ageing (see above). Brain volume loss and ventricular enlargement are highly correlated, but this relationship may theoretically be altered by treatments or by pathology (e.g. obstructive hydrocephalus). An extensive discussion of the literature relating to rates of longitudinally derived global brain atrophy rates in AD is given later in this thesis.

2.3.3.2 Regional atrophy

As the pathological progression of AD is known to start in medial temporal lobe structures, many authors have assessed longitudinal atrophy of MTL structures and brain substructures, and particularly the hippocampus as diagnostic or progression markers of AD. Table 2-1 provides a summary of some of the volumetric MRI studies assessing hippocampal atrophy rates.
Table 2-1. Annual longitudinal hippocampal atrophy rates in AD and controls.

<table>
<thead>
<tr>
<th>Author</th>
<th>Year</th>
<th>Structure</th>
<th>Mean (SD) Rate(%)/yr in AD</th>
<th>Mean (SD) Rate(%)/yr in controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaye (Kaye et al, 1997)</td>
<td>1997</td>
<td>Hippo</td>
<td>2.3 (2.0)*</td>
<td>2.1 (3.5)</td>
</tr>
<tr>
<td>Jack (Jack et al, 1998)</td>
<td>1998</td>
<td>Hippo</td>
<td>3.98 (1.9)</td>
<td>1.55 (1.4)</td>
</tr>
<tr>
<td>Laakso (Laakso et al, 2000)</td>
<td>2000</td>
<td>Left Hippo</td>
<td>7.2 (20.1)</td>
<td>3.6 (15.1)</td>
</tr>
<tr>
<td>Laakso (Laakso et al, 2000)</td>
<td>2000</td>
<td>Right Hippo</td>
<td>6.9 (18.2)</td>
<td>1.6 (15.1)</td>
</tr>
<tr>
<td>Jack (Jack et al, 2000)</td>
<td>2000</td>
<td>Hippo</td>
<td>3.5 (1.8)</td>
<td>1.7 (0.9)</td>
</tr>
<tr>
<td>Jack (Jack et al, 2003)</td>
<td>2003</td>
<td>Hippo</td>
<td>4.9 (2.1)**</td>
<td>-</td>
</tr>
</tbody>
</table>

*: Initially asymptomatic AD; **: median, estimated SD

As can be seen from the table, although excess atrophy compared to controls is seen in all studies, there is large variability in the actual rate of calculated hippocampal atrophy between studies, and in some cases there is also significant within-study variation (reflected by a large standard deviation). The former is likely to be due to a combination of measurement error and differences in measurement protocols; the latter to be due to inter- and intra-rater variability in measuring the hippocampus, as well as the possibility that the hippocampus might undergo different rates of atrophy at different disease stages. In line with the histopathological findings in AD, studies in familial AD have demonstrated that excess hippocampal and entorhinal cortex atrophy is probably occurring years prior to the clinical manifestation of the disease (Fox et al, 1996a; Schott et al, 2003), and excess medial temporal lobe atrophy can predict which patients will proceed from a diagnosis of MCI to AD (Jack et al, 2000). Conversely it may be that the rate of medial temporal atrophy decreases later in the disease process (Silbert et al, 2003; Scahill et al, 2002), perhaps when the hippocampus has little more volume to lose. Although entorhinal cortex atrophy has
been reported to undergo both earlier and higher rates of atrophy than the hippocampus, most authors conclude that hippocampal measures (which are less prone to measurement error) are likely to be overall a more practical marker of disease progression (Xu et al, 2000; Jack et al, 2004; Schott et al, 2003).

2.3.3.3 Conclusions

The results of volumetric studies allow some conclusions concerning atrophy in AD to be drawn. AD is associated with early medial temporal lobe atrophy that is accompanied by increased rates of whole brain atrophy. The rate of MTL atrophy is likely to be higher than that of whole brain atrophy particularly in early AD, but is likely to show more variability due to measurement errors associated with manual outlining, and possibly variability in rate over the course of the disease. Rates of either medial temporal lobe or whole brain atrophy are increased in AD compared to age-matched controls; this excess atrophy is an inevitable consequence of the progression of the disease, and therefore provide a signal that could be used as a surrogate marker of AD progression for treatment trials: if a drug which improved cognition also was to reduce the rate of atrophy in the treated compared to placebo group, this may, with the appropriate trial design and with some important caveats, suggest a disease modifying rather than merely symptomatic effect.
3 Atrophy as a surrogate marker in AD trials

At the time of writing, all available treatments for AD are licensed on the basis of having symptomatic rather than disease modifying effects. Drug licences are granted on the basis of results of randomised placebo controlled clinical trials, which are the standard methodology applied to assess the efficacy and safety of methods used for the treatment and prevention of disease. Trial design and statistical methodologies have evolved to address issues such as control of patient variability, bias, compliance and medical ethics. Such clinical trials require considerable time for completion and are extremely expensive to run. The possibility that new treatments for AD will have disease-modifying properties brings new challenges to clinical trials, which must be able to determine symptomatic from disease modifying effects; there will also be an urgent need to establish efficacy and safety in the shortest possible timeframe to allow patients to be treated as soon as possible. MRI based measures have been proposed as surrogate markers of AD progression that could be used in clinical AD trials both to help distinguish disease-modifying from symptomatic effects, but also to reduce the length (and thereby cost) of such studies (Matthews et al, 2003).

3.1 Current regulatory practice for AD drugs

The Food and Drugs Administration (FDA) is the legislative body responsible for drug licensing in the USA, and has considerable influence over how drugs are licensed elsewhere in the world. Based on criteria released in 1990 (Leber, 1990), the FDA requires the following before licensing a drug for dementia:

1. Demonstration of efficacy should be shown in at least two randomized control trials that are each of at least 3-6 months duration;
2. Efficacy should be established using the Alzheimer’s disease assessment scale-cognitive (ADAS-Cog) (Rosen et al, 1984) (see 1.2.6.1);

and either:

3. A clinician’s overall impression of how the patient’s cognitive, behavioural and function has changed over the course of the study. The most commonly used test is the Clinician Interview Based Impression of Change-plus – CIBIC-plus, which is assessed by a blinded clinician on a scale from 1 (marked
improvement) to 7 (marked worsening); a rating of 4 denotes no change (Schneider et al, 1997).

or:

4. A measure of activities of daily living

The FDA also requires that statistically significant superiority ($p<0.05$) of the active drug over placebo be demonstrated for each of the two primary outcome measures (Mani, 2004). In the case of all the cholinesterase inhibitors licensed to date, the ADAS-Cog and CIBIC-plus have been the primary efficacy measures in key studies upon which approval by the FDA has been granted.

### 3.2 Potential problems with current outcome measures

Whilst the utility of neuropsychological tests and rating scales in detecting worsening of symptoms has been demonstrated in longitudinal studies as well as in trials of investigational drugs for symptomatic palliation of AD, the power of these tests to quantify drug effects especially with respect to disease progression is still unclear. The ADAS-Cog, whilst proving useful in assessing disease severity and drug efficacy is not without its problems. ADAS-Cog ratings may depend on age and education (Matthews et al, 2003); and when used as an outcome measure in pharmaceutical trials, changes in ADAS-Cog scores may be attributed to symptomatic effects of investigational drugs or to effects on the underlying rate of disease progression. For example, a symptomatic effect was demonstrated in studies of individuals treated with cholinesterase inhibitors whose improvements in ADAS-Cog scores during treatment declined to placebo-treated levels after drug withdrawal (Raskind et al, 2000; Rogers et al, 1998). If, conversely, patients had not returned to placebo-treated levels post-washout, an effect on disease progression would have been suggested (Whitehouse et al, 1998). Two theoretical study designs have been proposed to assess this distinction prospectively. Both designs apply to studies that are randomized, double-blind, placebo-controlled and parallel-arm throughout, and each proposed design consists of two study segments.

In a randomized withdrawal design, patients are initially randomized to either active drug or placebo. This segment is then allowed to continue for a sufficient duration to
allow the active drug to demonstrate efficacy in relation to placebo. At the beginning of the second study segment, those randomized to active drug in the initial phase are further randomized to either continue active drug or receive placebo. The second study segment then continues for an appropriate period. If at the end of the second segment, those receiving placebo, in that phase only, maintained their difference from those who received placebo through both segments (Fig 3-1, scenario A), a disease-modifying effect would be assumed. On the other hand should those receiving placebo in the second segment deteriorate to the level of those who received placebo throughout (Fig 3-2, scenario B), a purely symptomatic effect would be inferred.

Figure 3-1. Randomized withdrawal placebo-controlled design studies.
A: result expected - disease-modifying drug; B: result expected - symptomatic drug

In a randomized start design, patients are again randomized to active drug or placebo in the initial segment. This segment is then allowed to continue for a sufficient duration to allow the active drug to demonstrate efficacy in relation to placebo. At the end of that period those who received placebo during the initial segment are rerandomized to receive active drug or placebo for the entire duration of the second segment. If at the end of the second segment the group which received placebo initially “catches up” with those who received active drug throughout a symptomatic effect is inferred (Fig 3-2, scenario A); on the other hand if a difference between the groups is maintained, the active drug is assumed to have a disease-modifying effect (Fig 3-2, scenario B),
Figure 3–2. Randomized start placebo-controlled design studies.
A: result expected - symptomatic drug; B: result expected - disease-modifying drug

However studies designed to investigate drugs intended to slow disease progression in either of these ways are large, lengthy, and complex (Whitehouse et al, 1998). Progression of disease usually occurs over months to years, and so trial designs would require staggered wash-out or wash-in of an investigational drug to ensure that effects of drug and placebo can be dissociated (Matthews et al, 2003). As a consequence, even proof-of-concept studies that are properly powered to show an effect on disease progression require hundreds of patients taking experimental drugs for extended periods. Furthermore, both these study designs can still be considered theoretical and have yet to be adequately assessed in a clinical trial setting; the appropriate durations of each segment, the frequency of assessments, and a number of analytical issues have yet to be definitively established; and to date, there is no instance when either of these approaches has been successfully applied in a clinical trial of a putative disease-modifying effect in AD (Mani, 2004). In an effort to reduce both the completion time (allowing drugs to be brought to the market more quickly) and required patient numbers (reducing the cost of trials and exposure to potential side-effects), surrogate markers of disease progression are increasingly being considered as alternative or additional outcome measures in clinical trials.
3.3 Surrogate markers

A widely quoted definition of a surrogate endpoint is that proposed by Temple (Temple, 1999).

“A surrogate endpoint for a clinical trial is a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions, or survives. Changes induced by therapy on a surrogate endpoint are expected to reflect changes in a clinically meaningful endpoint.”

According to Fleming and DeMets (1996), the properties of an ideal surrogate endpoint should be as follows:

1. A proposed surrogate endpoint must not merely be a correlate of the true clinical outcome.

2. The effect of an intervention on a valid surrogate endpoint must reliably predict the effect of the clinical outcome of interest.

3. The treatment effect on the clinical outcome should be explained by its effect on the surrogate marker.

These concepts are illustrated in Figure 3-3.

Figure 3–3. Diagram to illustrate ideal properties of a surrogate marker.

The disease inevitably leads to change in the surrogate marker which inevitably leads to the true clinical outcome. The intervention affects both the surrogate and thus inevitably the true outcome.
3.3.1 Successful surrogate markers

To date, few (if any) surrogate markers have been validated in any clinical setting, and there is considerable debate within the scientific community as to how surrogate markers should be evaluated (Buyse et al, 2000). An example of a surrogate in another field of medicine is CD4 count as a marker of human immunodeficiency virus (HIV) infection. The HIV virus attacks and destroys CD4 cells. Reduction of CD4 count leads to immune deficiency which leads to increased mortality in patients infected with HIV. Drugs designed to decrease mortality by impacting on HIV should by definition increase CD4 count; this increased CD4 count should decrease mortality rate. Thus CD4 count is proposed (and used) as a surrogate marker for mortality (the true outcome measures) in HIV treatment trials. However, even in this case there is controversy as to how good as surrogate CD4 count is, as not all trials have shown a consistent relationship between improved CD4 counts and decreased mortality (for a critical review see (Fleming and DeMets, 1996); from a practical perspective, in HIV trials CD4 count is usually used in combination with other markers of the disease (Jordan et al, 2002), and it is likely that in all trials, rather than reliance on one surrogate, combinations of outcome measures will be used.

3.3.2 Failed surrogates

Many proposed surrogate markers have in practice failed. A well known example of a failed surrogate end-point is that of ventricular arrhythmias as a surrogate marker for mortality after myocardial infarction. In the cardiac arrhythmia suppression trial (CAST), three drugs (encainide, flecainide and moricizine) were approved for marketing on the basis that they suppressed ventricular arrhythmias after myocardial infarction, ventricular arrhythmia being widely assumed to be a major cause of mortality post myocardial infarction (Anon., 1992; Anon., 1989). Subsequently formal studies using randomized double-blind, placebo-controlled trials, demonstrated that mortality was increased in patients on these drugs, leading to licenses for this indication being withdrawn.
3.3.2.1 Why do surrogates fail?

Fleming and DeMets (Fleming and DeMets, 1996) suggest a number of reasons why surrogate end-points may fail.

1. Although a surrogate may be a correlate of the disease process, its progression may involve a different pathophysiological process; thus intervention that affects the surrogate may not influence the true clinical outcome (Fig. 3-4; A);

2. A disease may progress through several causal pathways, not all of which are mediated through the surrogate. Thus an intervention affecting the surrogate may not fully predict the effect on the true clinical outcome (Fig. 3-4; B);

3. Conversely, an intervention might affect a pathway leading to the true clinical outcome without impacting on the surrogate (Fig. 3-4; C); or

4. The intervention might, through unintended mechanisms, affect the true clinical outcome independent of the disease process (Fig. 3-4; D).

Figure 3-4. Diagram to illustrate reasons surrogate markers may fail. Adapted from Fleming and DeMets (1996).
3.3.2.2 Regulatory views on surrogate markers

Despite the potential problems associated with surrogate markers of disease progression, the FDA is able to approve drugs using surrogate markers under its accelerated approval regulations (Anon., 2001). These regulations apply to drugs studied in “serious or life-threatening illnesses” which produce “meaningful therapeutic benefit to patients over existing treatments”. First instituted in 1992 at a time when there was an urgent need to develop effective drugs for the treatment of acquired immune deficiency syndrome (AIDS), it has been proposed that disease-modifying drugs for AD could fall within this remit.

Under these regulations, the FDA is permitted to grant marketing approval to drugs based on a surrogate endpoint that appears “reasonably likely” to predict clinical benefit based on “epidemiological, therapeutic, pathophysiologic, or other evidence”, with actual clinical benefit being confirmed after approval. The subsequent FDA modernization act signed into US law in 1997 allow for “approval of a fast track product upon a determination that the product has an effect on a surrogate marker that is reasonably likely to predict clinical benefit”, with the caveat that appropriate post-marketing studies be carried out to validate the surrogate endpoint (Anon., 1997b).

3.3.3 Imaging as a surrogate marker of AD progression

Since the first proposals that imaging measures of atrophy might be useful in therapeutic trials in AD (Smith and Jobst, 1996), using imaging (and particularly MRI) based measures of atrophy in this way has been the subject of growing interest within the scientific and pharmaceutical industries (Kaye, 2000; Matthews et al, 2003), and has been the subject of a number of natural history studies (Jack et al, 2003; Fox et al, 2000; Kaye, 2000; Bradley et al, 2002). Indeed many drug studies including the failed Milameline (Jack et al, 2003) and interrupted amyloid vaccination (AN1792-QS21) studies have already used MRI based measures of atrophy as a surrogate marker of AD progression.

Interest in using atrophy as an outcome measure prompted the FDA to convene a meeting of its joint advisory committee to discuss “Issues Related to the Role of Brain Imaging as an Outcome Measure in Phase III Trials of Putative Drugs for Alzheimer’s
Disease” in November 2002. In 2004, a senior spokesman from the FDA concluded that “Brain imaging may have the potential to be used as surrogate markers to support a regulatory claim that a drug has a disease-modifying effect in AD. However, prior to the acceptance of one or more such modalities for that purpose a number of questions need to be addressed.” (Mani, 2004). To answer these questions, he concluded that questions surrounding the use of imaging as a surrogate could be addressed by (1) comparing the imaging surrogate marker with the desired clinical outcome across multiple drugs and clinical trials; and (2) explaining the biological basis for the intervention-induced change in the surrogate marker.

In the absence of a proven disease modifying drug, it is not possible to determine the effect of such drugs on disease progression as measured using MRI. However, increasing evidence is now accumulating from natural history studies and treatment trials further supporting the case for the use of MRI measures of progression as surrogate markers in AD.

### 3.3.3.1 Biological plausibility

As discussed in the previous chapter, numerous studies have confirmed that excess cerebral atrophy occurs prior to the onset of symptoms in both familial (Fox et al, 1996a; Fox et al, 1999b; Schott et al, 2003) and sporadic AD (Rusinek et al, 2003), and is an inevitable consequence of the progression of AD (For review see (Fox and Schott, 2004)). The pattern of progressive atrophy in AD demonstrated using an unbiased assessment of longitudinal MRI (Scahill et al, 2002) has been shown to correlate with the histopathological progression of the disease closely (for review see (Smith, 2002)). Silbert et al have recently demonstrated evidence that MRI measures of atrophy in life are predictive of changes seen at post-mortem (Silbert et al, 2003). These authors found that total brain volume change over time was related to the accumulation of cortical neurofibrillary tangles and that the rate of ventricular CSF volume increase was related to both cortical neurofibrillary tangles and senile plaques. There is also evidence that increased rates of atrophy have clinical correlates in vivo. Thus Fox et al found a correlation between change in whole brain volume and change in MMSE (Fox et al, 1999a). These studies, and many others, consistently demonstrate that excess cerebral atrophy is an inevitable consequence of AD, at least
in untreated patients, which has clinical correlates, and which can reliably be quantified using serial MRI.

### 3.3.3.2 Feasibility

As previously described, a number of studies have quantified rates of whole brain or regional cerebral atrophy derived from serial scans over periods usually in excess of one year. Rates of atrophy derived in this way have been used to estimate sample sizes that would be required to power a drug trial. Fox et al described the use of standard power formulae to determine sample sizes that would be needed in each arm of a placebo controlled trial to provide 90% power at a 5% significance level (paired) to detect a 20% slowing of atrophy (Fox et al, 2000), where:

\[
\text{Sample size per treatment arm} = (u + v)^2 \times \left( \sigma_1^2 + \sigma_2^2 \right) / (\mu_1 - \mu_2)^2
\]

Where \( u = 1.28 \) (to provide 90% power); \( v = 1.96 \) to test at the 5% level, \( \sigma_1 \) and \( \sigma_2 \) are the standard deviations of the placebo and test group respectively, and \( \mu_1 \) and \( \mu_2 \), are the mean of the placebo and test group respectively.

Sample sizes can then be corrected to account for the effects of normal ageing (i.e. assuming that a true 20% reduction of disease-related atrophy is 20% of the difference between the rate of atrophy in the patient group and the controls) and for the estimated effects of drop out or unusable scans. This formula can be used to estimate sample sizes that would be required to power studies based on previously reported rates of regional and whole brain atrophy. Sample sizes based on hippocampal and whole brain atrophy rates from a variety of studies are shown in Table 3-1. These results suggest that the samples sizes that would be required in placebo controlled studies at one year are feasible using MRI measures as an outcome. From these limited studies, there appears to be little benefit in using manual hippocampal measures or global brain atrophy calculated using automated techniques; this has been further demonstrated in a recent head-to-head study, which showed that there was little difference in the numbers required per treatment arm whether hippocampal or whole brain measures were used (Jack et al, 2004). In one of the few large prospective,
multi-centre drug studies (halted due to lack of efficacy) using cerebral atrophy as an outcome measure, rates of hippocampal atrophy and change in ADAS-Cog were both assessed, and thus results can be directly compared (Jack et al, 2003). Power calculations under the same assumptions as those used in Table 3-1 show that using the ADAS-Cog, approximately 4000 patients would be required per treatment arm to detect 20% slowing over a one-year period; this number was 263 per arm when hippocampal volumes were used.

Table 3-1. Sample size estimates based on reported atrophy rates in AD.
Estimated sample sizes required per treatment arm to power a study to detect 20% atrophy, with 90% power at the 5% level (paired) not accounting for normal ageing or dropout.

<table>
<thead>
<tr>
<th>Author</th>
<th>Structure</th>
<th>Mean (SD) Rate/yr</th>
<th>Number in study</th>
<th>Mean follow-up (months)</th>
<th>Sample size estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack (Jack et al, 1998)</td>
<td>Hippo</td>
<td>3.98 (1.9)</td>
<td>24</td>
<td>22.7</td>
<td>120</td>
</tr>
<tr>
<td>Jack (Jack et al, 2000)</td>
<td>Hippo</td>
<td>3.5 (1.8)</td>
<td>18</td>
<td>34.8</td>
<td>139</td>
</tr>
<tr>
<td>Fox (Fox et al, 1999a)</td>
<td>Brain</td>
<td>2.4 (1.4)</td>
<td>29</td>
<td>21.6</td>
<td>179</td>
</tr>
<tr>
<td>Fox (Fox et al, 2000)</td>
<td>Brain</td>
<td>2.37 (1.1)</td>
<td>18</td>
<td>11.0</td>
<td>113</td>
</tr>
<tr>
<td>O'Brien (O'Brien et al, 2001)</td>
<td>Brain</td>
<td>2.0 (0.9)</td>
<td>9</td>
<td>~12</td>
<td>106</td>
</tr>
<tr>
<td>Bradley (Bradley et al, 2002)</td>
<td>Brain</td>
<td>2.14 (0.5)</td>
<td>5</td>
<td>2.5-7</td>
<td>29</td>
</tr>
<tr>
<td>Wang (Wang et al, 2002)</td>
<td>Brain</td>
<td>2.4 (1.2)</td>
<td>14</td>
<td>12</td>
<td>131</td>
</tr>
</tbody>
</table>
These results suggest that in terms of numbers required to power a study, measures of brain atrophy are feasible outcome measures in clinical trials of AD. However, many of the studies to date have investigated small numbers of patients, often in retrospective studies. Further validation is therefore required in larger groups of elderly patients with sporadic AD, followed prospectively.

3.3.4 Conclusions

The use of MRI based measures of atrophy that serve as surrogate markers for disease progression could provide complementary information or advantages over behavioral testing like the ADAS-Cog for several reasons. Imaging is likely to be more sensitive and specific in detecting change, making it possible to power studies using fewer subjects and/or shorter duration of exposure, as demonstrated by Jack et al (Jack et al, 2003). Unlike neuropsychological or behavioural scores, measures of atrophy are linear over short periods, and much less susceptible to ceiling or floor effects, bias or variability. As discussed above, as progressive excess atrophy appears to be an inevitable feature of the disease process closely allied to the underlying pathology of the disease, it is biologically plausible that progressive atrophy is a valid surrogate marker for true disease progression. Thus a trial demonstrating that a drug was produced a reduction in atrophy compared to placebo might provide evidence of a disease-modifying effect, in addition to any cognitive and behavioural benefits demonstrated using test such as the ADAS-Cog or CIBIC-plus.

3.3.5 Questions

"The use of serial volumetric MRI measures in studies of drugs that have the potential for altering the rate of underlying disease progression in AD seems warranted" (Matthews et al, 2003). If cerebral atrophy is to be used as an outcome measure in AD trials, a number of issues relating to the design of such studies remain in question.

It would clearly be advantageous to run drug studies over as short an interval as possible. To date, only one study (of 5 patients) has assessed atrophy at intervals shorter than one year (Bradley et al, 2002). It is not known whether prospective drug studies using MRI-based outcome measures are feasible over shorter intervals. The
precise rate of calculated cerebral atrophy in AD varies from study-to-study. An accurate understanding of the mean and variance of cerebral atrophy rate in AD and normal ageing could allow more accurate predictions of sample sizes required to power two time-point clinical studies which are critically dependent on the variance of any measures. There are theoretical advantages in combining multiple scan intervals together to decrease inter-individual variance in atrophy rates and therefore to reduce sample sizes required in studies. An accurate understanding of the sources of variance (i.e. between-subject, within-subject, and measurement error) could allow for mathematical modelling to explain the rates of atrophy in AD, and thus for such scenarios to be critically evaluated. Further information concerning the correlates of atrophy would strengthen the case for atrophy being a true surrogate marker of AD progression, and an understanding of the predictors of cerebral atrophy could help both our understanding of AD pathology and possibly allow stratification of patients in treatment trials.

The aim of this study was to evaluate these questions critically, using global brain atrophy as an outcome measure. Global brain atrophy was chosen over regional (especially hippocampal) atrophy for a number of reasons. Whilst absolute measures of rate of hippocampal change are generally greater than rates of whole brain atrophy, the former are associated with greater measurement errors, arbitrary decisions regarding boundaries are required, and hippocampal change may not be a valid biomarkers of the disease process towards the end of the disease (Silbert et al, 2003); the latter may also account for the finding in a recent study, that whole brain atrophy correlated better with cognitive change than did hippocampal change (Jack et al, 2004). Furthermore hippocampal outlining is time consuming, and making calculations of atrophy over large number of intervals (as required in this study design), impractical. Although global brain atrophy is proportionally of a smaller magnitude than some regional changes, it can be measured using automated tools (such as the BBSI), and is thus quicker, more robust and more practical for studies involving many hundreds of scans.
The specific aims of this study were to assess whether:

1. Patients with AD can tolerate trials involving multiple short scans, and whether strategies can be developed to maintain patients in multi-time point trials and obtain good quality MR images;

2. Nine-degrees-of-freedom registration is a valid method for correcting for voxel drift, ensuring that serial scans can be accurately compared without “correcting away” atrophy;

3. Altering the BBSI parameters materially influences calculations of global atrophy, and whether rates of atrophy calculated in this way can correctly classify patients from controls.

4. It is feasible to power drugs studies in AD using global brain atrophy at intervals of under one year;

5. There are advantages in using manual (ventricular volume) measures of global brain atrophy over automated measures (BBSI) in short two time-point studies;

6. It is possible to model atrophy from serial scans mathematically and thus to determine the mean, variance and sources of variance in atrophy rates in cohorts of patients with AD and controls;

7. There are material advantages in combining atrophy measures from more than two scanning points to reduce variability in atrophy rates and therefore of required sample sizes, and how best to design studies involving multiple scans;

8. Significant correlates of change and predictors of increased atrophy can be determined from such a model; and

9. Useful information regarding the patterns of regional atrophy can be obtained from serial scans in late onset AD using the combination of fluid registration and SPM.
4 Methods Overview

4.1 Subjects

Subjects for the studies described in this thesis were recruited as part of the MIRIAD (Minimal Interval Resonance Imaging in Alzheimer's disease and normal ageing) study. This proof-of-concept study was designed to determine the shortest interval required to detect differences between AD subjects and normal controls using measures from serial volumetric MRI. All patients were recruited from the Specialist Cognitive Disorders Clinic at the National Hospital for Neurology or from the memory clinic at the Chelsea and Westminster Hospital.

4.1.1 Prior Clinical Assessment

Prior to recruitment to the study, all patients had undergone full clinical assessments as part of their clinical care, including a full history taken together with a close informant and a comprehensive general medical and neurological examination. In addition, patients were investigated with: detailed neuropsychological testing; standard screening blood tests to exclude other treatable causes of cognitive problems such as impaired renal or liver function, B12 and thyroid function; EEG to exclude seizures, or identify patterns indicative of a particular type of dementia; and neuroimaging using MRI to exclude tumours and subdural haematoma. All patients had a prior diagnosis of AD which had been explained to them and their carer.

4.1.2 Case selection: exclusion and inclusion criteria

The aim of the study was to identify ~50 patients with mild-moderate AD, and ~20 appropriate age-matched controls. Inclusion criteria for the patients included: age over 55; diagnosis of probable AD according to NINCDS-ADRDA criteria (McKhann et al, 1984); patient and carer willing to participate and able to give informed consent; MMSE score >12 (Folstein et al, 1975), and CDR <3 indicating mild-moderate dementia (Morris, 1993). Inclusion Criteria for the controls subjects included: age, sex and education matched individuals willing to participate and able to give informed consent; age >55 years; no history of significant cognitive decline, head injury, stroke or psychiatric illness; MMSE > 26. Where possible, controls who were spouses of
patients participating in the study were recruited both for practical reasons and to aid accurate matching of the groups in terms of background, age and education. Exclusion criteria for either patients or controls included any degenerative neurological disease (other than AD for the patients), inability to tolerate MRI scanning, or any contraindication to scanning.

4.1.3 Consent and ethical considerations
The study was approved by the local research ethics committees at the National Hospital for Neurology and Neurosurgery, Queen Square. Informed written consent to the study was obtained from all study subjects and their appropriate legal representative. In addition specific written consent was obtained to undergo research MR scanning. Individuals were not scanned if they had a contraindication to magnetic resonance imaging such as a cardiac pacemaker or metallic implant or had a history of claustrophobia. Travelling expenses were refunded to study subjects but they received no other payments or inducements. All subjects were informed that participation in the study was entirely voluntary and that they could at any time withdraw from the study, and that withdrawal would not in any way influence any subsequent clinical care. Additionally a missed assessment would not preclude them from rejoining the study at a later date. All individuals were made aware that the data collected as a part of the study were confidential and for research purposes. It was emphasised that the study would not provide any specific results on an individual basis. The one exception to this rule was that if the MR scan revealed an unrelated but clinically important structural lesion such as a cerebral neoplasm they and their general practitioner would be informed of the scan results. All subjects signed consent for their data to be stored under the Data Protection Act.

4.2 Study design
An overview of the study design is given in Table 4-1.
Table 4-1. Study design

<table>
<thead>
<tr>
<th>Visit (weeks)</th>
<th>Interval (weeks)</th>
<th>Volumetric MRI</th>
<th>Clinical assessment</th>
<th>Neuropsychology</th>
<th>MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>x2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>x1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>x2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14</td>
<td>x1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>x1</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>36</td>
<td>x2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>52</td>
<td>x1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>~52-56</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 4.2.1 Clinical assessment

Details of the clinical assessments are given in chapter 5. Briefly, all subjects were assessed at visits 0 and 8 to ensure fulfilment of inclusion criteria, with no exclusion criteria. At the baseline assessment (visit 0) a clinical assessment, physical examination were performed, and a baseline CDR, MMSE and modified Hachinski ischemic score (Appendix 5) were administered (Rosen et al, 1980).

### 4.2.2 Neuropsychological assessments

A short comprehensive neuropsychology battery was administered to all subjects at the start of the study, and as close as possible to the final scan. The battery was chosen to be applicable to patients with mild-moderate AD and to age-matched controls, and to be completed in less than 90 minutes. Details of the neuropsychology battery are given in chapter 5. In brief, assessments were made of estimated premorbid IQ, and tests were administered to assess the following domains: verbal memory; visual
memory; vocabulary; fluid intelligence/reasoning; arithmetic; naming; and visuoperceptual skills.

4.3 Magnetic Resonance Imaging

4.3.1 Acquisition

Prior to the commencement of the study, pilot work had been completed to determine an appropriate MRI protocol based on a 3D T1-weighted gradient echo volumetric sequence. This involved evaluating candidate acquisition protocols in normal subjects and then patients with AD in order to select the most appropriate protocol for use in the larger study. Criteria for the scanning protocol included: minimum acquisition time to maximize patient compliance; high signal to noise ratio (SNR); minimum distortion, shading and susceptibility artefact; high contrast between grey matter and cerebrospinal fluid (CSF) and between grey matter and white matter, with the former being more important; anatomical accuracy; ease of outlining using the MIDAS semi-automated segmentation tool (see chapter 4.3.2); and no adverse effect on the BBSI.

All scanning was performed on the same 1.5T Signa Unit (GE Medical Systems, Milwaukee). T1-weighted volumetric images were obtained using a spoiled fast GRASS sequence technique with a 24-cm field of view and 256 x 256 matrix to provide 124 contiguous 1.5-mm-thick slices in the coronal plane. The scan acquisition parameters determined during the piloting stage were as follows: repetition time = 15ms; echo time = 5.4 ms; flip angle = 15°; inversion time = 650ms. The length of time to acquire a volumetric scan was under 10 minutes.

Details of methods used to aid subject compliance are given in chapters 6 and 7. In brief, subjects were reminded to remain as still as possible during the scanning and were given earplugs to wear. They were positioned supine on the scanner table with their head supported by foam wedges either side of the head. No restraint or sedation was used. Spouses or carers were permitted to remain in the scanner suite if subjects were anxious, and encouragement and explanation was provided during the scan as needed.
4.3.2 Post processing

4.3.2.1 Inhomogeneity correction and whole brain segmentation

After acquisition, digitised images were transferred to a Sun workstation (Sun Microsystems Inc., Mountain View, CA) for analysis. In order to counter potential inhomogeneity (chapter 2.1.2) all images were biased corrected using the N3 correction algorithm (Sled et al, 1998). Further image processing was performed within the Medical Information Display and Analysis System (MIDAS) software package, developed by Freeborough et al (Freeborough et al, 1996; Freeborough et al, 1997; Freeborough et al, 1997). This software allows for simultaneous multiplanar display and manipulation of 3D data. Using the N3 corrected images, each scan was segmented (scalped) using the method described by Freeborough (Freeborough et al, 1997). This is a semi-automated algorithm which first identifies voxels within the brain using interactive thresholding. Brain tissue is then isolated from surrounding tissue such as scalp and dura using a series of erosions and dilations (see Appendix 6).

4.3.2.2 Rigid-body registration and calculation of atrophy

To allow accurate comparisons between serially acquired scans, rigid-body registration was performed using a 9dof registration (Freeborough et al, 1996), with image reconstruction using chirp-z Fourier domain interpolation implemented within the AIR 5.2 package as described by Woods et al (Woods et al, 1992).

Measurement of ventricular volumes was performed on registered scans using the MIDAS package, which allows manual segmentation of regions of interest using a mouse-driven cursor. The simultaneous display of orthogonal views allows the operator to outline the structure in the coronal view whilst the segmentation is updated in real time in the sagittal or axial view. This aids in decisions about where boundaries should be defined. Details of the ventricular segmentation protocol are described in chapter 10 and Appendix 7.
Calculation of whole brain atrophy from the registered scan pairs was performed using the brain boundary shift integral (BBSI) (Freeborough and Fox, 1997). Details of the BBSI algorithm, and optimization of the technique for this study are described in chapter 12.

4.3.2.3 Fluid registration and unbiased assessment of regional atrophy

A previously described unbiased technique was used to determine significant areas of contraction and expansion occurring during the course of the disease (section 2.3.1.3) (Scahill et al, 2002). This process involves first assessing areas of expansion or contraction occurring over time in an individual using fluid registration performed within the MIDAS software package, using previously outlined parameters (Freeborough and Fox, 1998; Crum et al, 2001). Fluid registered (within-individual) scans thus derived can then be compared between groups of individuals using Statistical Parametric Mapping (SPM99) (Wellcome Department of Cognitive Neurology, London) running on a Matlab platform (MathWorks, Natick, Mass., USA). Further details of the methodology are given in chapter 15.

4.4 ApoE genotyping

Specific consent was obtained from patients to undertake ApoE genotyping, which was performed in the MRC Prion Unit, University London. Twenty ml of blood was collected by venepuncture. DNA was extracted using standard techniques and ApoE genotyping was determined using a standard one-stage polymerase chain reaction (Wenham et al, 1991).

4.5 Data management

To allow accurate management and analysis of the large amount of data produced by this study, a bespoke database was designed within the Access software package, running on the Windows 2000 operating system. Scan and ventricular volumes were automatically outputted from the MIDAS package into the database to avoid data entry errors. The database was password protected and stored on a secure system in accordance with the Data Protection Act, and with the consent of the Caldicott Guardian at the National Hospital for Neurology and Neurosurgery (UCLH), London.
4.6 Statistics

STATA version 8 (Stata Corporation, College Station, Texas, 2003) was used to perform standard parametric and non-parametric tests and were used to investigate basic linear regression and test group differences.

The multi-level model designed to analyse this data-set (Frost et al, 2004) and extended to determine the dependency of atrophy on covariates is described in chapters 13, 14 and 15. The model was implemented in SAS version 8.2 (SAS Institute Inc., Cary, North Carolina).

Power calculations were performed using standard formulae (Armitage and Berry, 1991), and were based on adaptations described by Fox et al (Fox et al, 2000). Unless stated, all calculations in this thesis were performed to estimate sample sizes required in each arm of a placebo-controlled trial to provide 90% power to detect a 20% reduction in atrophy at a 5% (paired) significance level. Where stated in the text, these estimates were corrected to account for the effect of atrophy associated with normal ageing.
5 Recruitment to a longitudinal imaging study

5.1 Introduction

In order to determine the potential utility of short-interval derived MRI atrophy as a surrogate marker of AD, it is first necessary to define a cohort of patients who are representative of patients who might enter a treatment trial. In order to recruit such a representative sample, the criteria for the study were designed to include patients with high clinical likelihood of having sporadic AD, whilst excluding patients with very atypical features, such as a history of young onset (<55), that might be suggestive of a familial form of the disease. The aim of the work described in this chapter was to recruit 50 patients with late-onset, sporadic AD, and 20 controls, and to define these cohorts as accurately as possible using a range of clinical measures, standardized batteries and neuropsychological testing.

5.2 Methods

Details of the inclusion and exclusion criteria are given in the methods section. All subjects were 55 or over; were willing to participate and able to give informed consent; and able to tolerate an MRI scan. Patients with AD had an MMSE between 12 and 26, and a CDR <3, and had probable AD based on the NINCDS-ADRDA criteria. Controls were 55 or over, without any history of significant cognitive decline, head injury, stroke or psychiatric illness. Subjects were excluded if they had any degenerative neurological disease (other than AD for the patients), or were unable to tolerate MRI scanning or had any contraindication to scanning.

Patients with AD were recruited either from the cognitive disorders clinic at the National Hospital for Neurology and Neurosurgery, or via the Memory clinic at the Chelsea and Westminster Hospital, carried a diagnosis of probable sporadic AD (McKhann et al, 1984), and conformed to the inclusion criteria. As, for ethical reasons, patients would be unlikely to stop cholinesterase inhibitors should they enter a treatment trial, patients on cholinesterase inhibitors were not excluded from entry to the study, although stability on or off treatment was desirable. Controls were healthy volunteers, and were generally the spouse or carer, aiding appropriate matching for age and education. Prior to recruitment, all subjects were given information sheets.
pertaining to the study. Subjects were then, with their permission, contacted by telephone. Those consenting to enter the study attended for a standardised baseline assessment.

5.2.1 Clinical assessments

At the initial visit, written informed consent was obtained for participation in the study, and to ensure compliance with the Data Protection Act. Subjects were interviewed using a semi-structured clinical assessment. Data collected included details of the onset, progression and current level of cognitive impairment, as well as a full drug and family history. Specific information collected for all subjects included: age, handedness, school-leaving age and education, smoking history, and occupation. For the patients an estimate of age at symptom onset was made from the history and past medical notes where available, and information regarding the use of cholinesterase inhibitors was collected. All patients underwent a physical examination, including measurement of recumbent blood pressure (using a standard mercury sphygmomanometer), and estimation of body mass index (BMI) calculated according to the formula: weight/(height)^2. All subjects were assessed using the MMSE (Folstein et al, 1975) and the modified Hachinski ischaemic score (score from 0 to 12), with vascular risk factors leading to an increased score (Rosen et al, 1980). Patients were assessed using the Clinical Dementia Rating (CDR) scale, calculated both as a global score (0.5, 1 or 2), and as a sum of boxes score (Morris, 1993) (Appendix 4). Results were analysed using a combination of descriptive, parametric (Student’s t-test), non parametric (Wilcoxon Rank sum test) and Chi-squared tests, as appropriate.

5.2.2 Neuropsychology

Each subject underwent testing at the start of the study (visit 0) and at ~one year (visit 8) using a standardized neuropsychological battery. The battery was designed to provide a comprehensive assessment of cognitive domains known to be affected in AD, to be implemented in less than ninety minutes, and to be applicable both to patients with mild-to-moderate AD and to aged-matched controls. The protocol comprised the following tests: Verbal memory: Easy recognition memory tests for words (RMT-w) (Clegg and Warrington, 1994); Visual memory: Easy recognition memory tests for faces (RMT-f) (Clegg and Warrington, 1994); Vocabulary:
Vocabulary scale - The Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999); Reasoning: Matrix reasoning – WASI (Wechsler, 1999); Arithmetic: Graded Difficulty Arithmetic Test (Jackson and Warrington, 1986); Naming: Graded Naming Test (McKenna and Warrington, 1983); Visuoperceptual skills: Silhouettes - Visual Object and Space Perception battery (Warrington and James, 1991); Premorbid IQ: National adult reading test (NART) (Nelson, 1982)

NART scores were converted to IQ scores, and were compared using t-tests. To allow a comparison between performances in different cognitive skills, other test results were converted to z-scores, where the z-score represents the number of standard deviations an individual's score is from the known mean age-related score for that test. z-scores thus determined were compared using parametric statistical methods (in this case unpaired Student’s t-tests). This method was applied to compare all tests, with the exception of the recognition tests for faces (RMT-f) and words (RMT-w). Choosing tests of memory to be applicable to patients with AD and normal controls presents difficulties; in order to produce a dynamic range across a wide range AD severity, the vast majority of cognitively intact subjects score at or near the ceiling on the test (25/25). Consequently the use of z-scores for this test is not appropriate, as all patients’ scores will be many orders of magnitude away from the mean. Thus to allow analysis of these tests, scores were categorised to seven grades assigned one of seven grade scores, according to the percentile score, as previously described (Newman et al, 1994) i.e <1%, Grade 0; 1-5% Grade 1; 5-10%, Grade 2; 10-25% Grade 3; 25-50%, Grade 4; 50-75% Grade 5; 75-100% Grade 6. Scores thus converted were compared between subjects using non-parametric (Wilcoxon rank sum test) statistics.

Two combined scores were calculated. The first, a composite cognitive score, was calculated as the average of the z-scores of the five tests (excluding the NART), where available. The second, a composite memory score was calculated as the average of the two grade scores.

5.2.3 ApoE genotyping

Patients with AD (but not controls) who gave additional consent, had blood taken for determination of Apolipoprotein E genotype as previously described.
5.3 Results

In total, 46 patients and 23 controls were recruited to the study. An additional 2 patients were assessed and did not take part. Of these, one was thought to have an alternate diagnosis (FTLD), and one was unable to tolerate a first MRI scan. Four controls were assessed and did not continue in the study, all unable to tolerate scanning, three because of claustrophobia, one due to kyphoscoliosis. Recruitment was achieved at the rate of approximately one subject per fortnight.

5.3.1 Clinical characteristics

The baseline clinical characteristics of each individual are shown in Appendix 8.

5.3.1.1 All subjects

Age
The average age at entry to the study was 69.4 (SD 7.1; range 55.7 – 85.9) for the patients, and 69.7 (SD 7.2; range 58.0 - 85.8) for the controls (no significant difference).

Sex
Of the 23 controls, 12 (52%) were male, and of the 46 patients, 19 (41%) were male (no significant difference).

Mini-mental state examination
In the patient group, the average MMSE (maximum score 30) was 19.2 (SD 4.0; range 12-26); for the controls it was 29.4 (SD 0.8; range 27-30), a highly significant difference ($p<0.0001$, t-test)

Handedness
There was no significant difference in the rate of left-handedness between the groups; 8.7% (4/46) of the patients were left-handed, compared with 13% (3/23) of the controls ($p=0.82$, Chi-squared test).
Blood pressure

Blood pressure was recorded for 22 controls and 42 patients. There was no significant difference in systolic, diastolic, or mean (diastolic + 1/3 (systolic + diastolic)) blood pressures between the patients and controls (Table 5-1).

Table 5-1. Blood pressure of patients with AD and controls.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Systolic BP mmHg</th>
<th>Diastolic BP mmHg</th>
<th>Mean BP mmHg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>43</td>
<td>143 (20)</td>
<td>84 (9)</td>
<td>103 (10)</td>
</tr>
<tr>
<td>Controls</td>
<td>22</td>
<td>144 (18)</td>
<td>84 (10)</td>
<td>103 (13)</td>
</tr>
</tbody>
</table>

Body mass index

Body mass index was recorded for 21 controls and 43 patients. For the controls the mean BMI was 26.0 (SD 3.8, range 24.3 – 27.8), and for the patients 23.9 (SD 3.3, range 22.9 – 24.8). Patients had significantly smaller BMIs than controls ($p=0.01$, Wilcoxon Rank Sum test).

Education

There was no difference in the number of years of school education, as estimated by the school leaving age, between the groups: controls mean 16.5 (SD 1.4) years, patients with AD mean 16.3 (SD 1.5) years. 30% (14/46) of the patients and 39% (9/23) of controls had a university degree (non-significant, $p=0.47$, Chi-squared test).

Smoking

Of the controls, 15/23 (65%) were life-long non-smokers. This was not significantly greater that in the AD cohort: 24/46 (52%) ($p=0.33$, Chi-squared test).
5.3.1.2 Patients

The following information was collected only for the patient group.

*Years of symptoms*

On average every patient was 4.6 years (SD 2.3) years into the clinical course of their disease, (range: 1.5 to 11 years).

*Cholinesterase inhibitors*

At the initial visit, 28/46 (61%) of patients were taking a cholinesterase inhibitor.

*Clinical Dementia Rating (CDR)*

According to CDR criteria, 7 patients had questionable dementia (CDR 0.5), 35 mild dementia (CDR 1), and 4 moderate dementia (CDR 2). The mean MMSE was, as expected, lower in those with higher CDR scores, falling from 23.3 in those with a CDR of 0.5, to 18.8 in those with CDR 1, to 15.3 in those with CDR 2. The sum of boxes score was significantly negatively correlated with the MMSE ($p<0.0001$).

*Modified Hachinski ischaemic score*

No AD patient had significant vascular risk. 81% had a score of 0; 12% had a score of 1; and 4.7% had a score of 2; and 2% (1 patient) had a score of 3 (maximum score 12).

5.3.2 Neuropsychology

At the baseline assessment significant differences were seen between patients and controls in all the following tests: Estimated premorbid IQ (NART), RMT-w, RMT-f, silhouettes, arithmetic, naming, vocabulary, matrix reasoning, and in the combined memory and cognitive scores. These are illustrated in Table 5-2.
Table 5-2. Baseline cognitive scores.
Memory score (RMT-w, RMT-f and composite memory scores) are presented as grade scores (maximum score = 6), and compared using the Wilcoxon signed rank test. Estimated premorbid functioning based on the NART is presented as a full scale predicted IQ. All other tests are presented as z-scores and compared using Student’s t-tests. (*: p<0.001, Wilcoxon signed rank test; †: p<0.001, t-test)

<table>
<thead>
<tr>
<th></th>
<th>AD Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>Estimated premorbid full scale IQ (NART)</td>
<td>40</td>
<td>105 (13)</td>
</tr>
<tr>
<td>RMT-words Grade score</td>
<td>43</td>
<td>1.23 (1.19)</td>
</tr>
<tr>
<td>RMT-faces Grade score</td>
<td>43</td>
<td>1.44 (1.71)</td>
</tr>
<tr>
<td>Composite memory grade score</td>
<td>43</td>
<td>1.34 (1.29)</td>
</tr>
<tr>
<td>Silhouettes z-score</td>
<td>42</td>
<td>-2.34 (1.5)</td>
</tr>
<tr>
<td>Arithmetic z-score</td>
<td>40</td>
<td>-2.03 (1.16)</td>
</tr>
<tr>
<td>Naming z-score</td>
<td>39</td>
<td>-2.61 (1.82)</td>
</tr>
<tr>
<td>Vocabulary z-score</td>
<td>42</td>
<td>-0.4 (1.5)</td>
</tr>
<tr>
<td>Matrix reasoning z-score</td>
<td>42</td>
<td>-1.19 (1.4)</td>
</tr>
<tr>
<td>Composite cognitive z-score</td>
<td>44</td>
<td>-1.7 (1.1)</td>
</tr>
</tbody>
</table>
5.3.3 ApoE genotype

Of the 46 patients, ApoE genotyping was performed on 42. 32/42 (76%) of cases had one or more E4 allele, and 10/42 had no E4 allele. 3/42 (7%) possessed an E2 allele. Based on the allele frequency of the general population (E2: 0.08; E3: 0.78; E4: 0.14 (Growdon, 1998)), the expected frequency of each genotype was calculated and compared with that found in this group (Table 5-3). Patients with AD were significantly more likely to have an E3E4 or E4E4 genotype, and significantly less likely to have an E3E3 genotype than expected from population based results.

Table 5-3. Expected and observed ApoE genotype vs mean age at onset.

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Expected frequency n (%)</th>
<th>Observed frequency n (%)</th>
<th>Mean age at onset (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2E3</td>
<td>5 (12%)</td>
<td>1 (2%)</td>
<td>55.0 (-)</td>
</tr>
<tr>
<td>E2E4</td>
<td>1 (2%)</td>
<td>2 (5%)</td>
<td>65.0 (9)</td>
</tr>
<tr>
<td>E3E3</td>
<td>26 (61%)</td>
<td>9 (21%) *</td>
<td>62.8 (7)</td>
</tr>
<tr>
<td>E3E4</td>
<td>9 (22%)</td>
<td>22 (52%) *</td>
<td>67.0 (7)</td>
</tr>
<tr>
<td>E4E4</td>
<td>1 (2%)</td>
<td>8 (19%) *</td>
<td>64.3 (3)</td>
</tr>
</tbody>
</table>

*: p<0.05, Chi² test

5.4 Discussion

These results confirm that it is possible to recruit well matched case-control cohorts for studies in AD. The patients and controls recruited had very similar ages, sex-ratios, handedness, blood pressures, smoking histories and levels of education. However as expected, there were clear differences in performances on tests of cognition. The patient group performed at a significantly lower level on both the screening MMSE, as well as on formal neuropsychological testing. It is notable that at least 80% of patients were able to complete any one of the chosen neuropsychological tests, demonstrating that it is possible to design and analyze relatively short comprehensive batteries that can be used by both patients with mild-moderate AD and aged matched...
controls. The results of both the MMSE and CDR were, as expected, correlated, and demonstrated that this cohort is heterogeneous in terms of disease stage, with patients ranging from very mildly to moderately impaired.

As previously discussed, a definitive diagnosis of Alzheimer's disease is not possible without histopathological confirmation. Diagnosis therefore relies on the use of diagnostic criteria, the most commonly used of which for AD is the NINCDS-ADRDA (McKhann et al, 1984). All patients recruited fulfilled the criteria for probable AD. These criteria however have limitations, and whilst sensitivity at detecting Alzheimer's disease is high, specificity is low (Hogervorst et al, 2000) (Varma et al, 1999). The results presented in this chapter provide some further indirect evidence that this cohort as a whole were likely to indeed be suffering from sporadic AD. No patient had a modified Hachinski ischemic score over 3; a cut-off over 4 has been demonstrated to improve the differentiation of AD from vascular dementia (Moroney et al, 1997). As well as profound memory impairments, the neuropsychological core feature of AD (for review see (Lambon Ralph et al, 2003)), the AD group were significantly impaired in other cognitive domains known to be impaired in Alzheimer’s disease including calculation (Martin et al, 2003) and visuoperceptual skills (silhouettes) (Binetti et al, 1996), tests sensitive to left and right parietal dysfunction respectively. The relative preservation of vocabulary in the patients is likely to be both a consequence of the patients having AD, where vocabulary may be relatively preserved until late in the disease (Wechsler, 1999; Camus et al, 2003), and the high education level of the cohort. However, vocabulary may not provide as good an estimate of premorbid IQ as the NART (Nelson, 1982; Crawford et al, 1988). The mean IQ for both groups was still in excess of 100, suggesting that, as expected, the NART is relatively immune to changes in mild-moderate AD (Nelson, 1982). Nonetheless the NART estimated premorbid IQ was significantly lower in the patients compared to the controls. This is likely to be due to the fact that the NART is a less reliable estimate of premorbid IQ as the disease progresses (Cockburn et al, 2000).

The smaller BMI in this AD cohort compared to controls is in keeping with previous reports of weight loss in dementia (Gillette-Guyonnet et al, 2000). There was no evidence in this small sample for blood pressure in patients to be either higher or
lower than that of controls. This is in keeping with previous reports; whilst midlife systolic hypertension is a risk factor for the development of AD (Kivipelto et al, 2001), by the time of diagnosis, the blood pressure is either normal or often low, although whether hypotension is a cause or effect of AD is not fully established (Davis et al, 2003).

A high percentage of the AD cohort had one or more ApoE4 alleles, and few had an E2 allele; indeed patients in this sample were significantly less likely to have the E3E3 genotype and significantly more likely to have either the E3E4 or E4E4 genotype than expected from control population studies, providing further supportive evidence that this cohort is likely to contain a high proportion of cases of AD (Growdon, 1998). Numerous studies have confirmed an increased risk of AD associated with ApoE4, and a decreased rate of AD associated with the presence of the ApoE2 allele (for review see (Higgins et al, 1997)); several studies have reported that this association may be particularly relevant in patients who develop AD between the ages of 65 and 75 (Frisoni et al, 1998), the modal age range in this study. The findings in this study of E4 allele positivity of 76% are broadly in keeping with the 66% found in a pathologically confirmed AD series (Weiner et al, 1999). However, perhaps due to the small sample size, there was no evidence that those patients with one or more E4 alleles had an earlier age of onset than those without, as has previously been reported (Growdon, 1998).

Despite the limitations inherent in a clinical diagnosis of AD, the features of this cohort appear compatible with those expected of a group of patients with AD who might enter a treatment trial. Having tolerated an initial MRI scan, these patients and the well-matched control group, were therefore entered into the prospective longitudinal study.
6 Longitudinal follow-up: retention and clinical change

6.1 Introduction

For short interval serial MRI scanning to be a useful tool in monitoring the progression of AD, it is necessary to demonstrate not only that the technique is valid, but also that it is practically feasible. To date, few studies have assessed whether serial imaging in AD is feasible at intervals of less than one year (Bradley et al, 2002), and questions have been raised as to whether patients with mild-to-moderate AD can tolerate multiple serial scanning. The aim of this work was to undertake a prospective scanning protocol involving subjects attending for a number of scans over carefully defined intervals ranging from two weeks to one year. The work undertaken for this chapter was performed (1) to assess the tolerability of multiple serial MRI scanning in AD, and methods of increasing retention; and (2) to assess measures of clinical and neuropsychological change occurring over the course of the study.

6.2 Methods

Following the baseline assessment, all subjects entering the study were requested to attend for seven scanning sessions over the following year. The scanning schedule is shown in Table 6-1. The scanning intervals ranged from two weeks to one year; in addition, whilst the trial was ongoing, selected subjects who had completed the one-year trial were invited to return for additional scans at 18/12 and 2 years. At three time-points subjects had two back-to-back scans, to allow assessments of the value of an additional scan at any one time-point on sample size estimates, and to see whether this would increase the number of scans available for analysis.
Table 6-1. Scanning schedule.

<table>
<thead>
<tr>
<th>Time (weeks)</th>
<th>Visit 0</th>
<th>Visit 1</th>
<th>Visit 2</th>
<th>Visit 3</th>
<th>Visit 4</th>
<th>Visit 5</th>
<th>Visit 6</th>
<th>Visit 7</th>
<th>Visit 8</th>
<th>Visit 9</th>
<th>Visit 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volumetric MRI</td>
<td>x2</td>
<td>x1</td>
<td>x2</td>
<td>x1</td>
<td>x2</td>
<td>x1</td>
<td>x1</td>
<td>x1</td>
<td>x1</td>
<td>x1</td>
<td>x1</td>
</tr>
<tr>
<td>MMSE</td>
<td>x1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: Optional visits, not initially scheduled

Each subject was given details of every scanning visit for the whole study (including both the date and time of each scan) at the start of the study (i.e. up to one year in advance). All scanning was scheduled for either a Saturday morning or Wednesday evening. Before each visit, the patient and/or carer was telephoned to remind them of their appointment, and subjects were sent taxis for travel to and from the scanner. Each visit was scheduled for 40 minutes. At every visit the subject and carer were met either by myself or a colleague to whom they had previously been introduced, and a brief safety checklist was completed (Appendix 9) to ensure there were no new contraindications to MRI scanning, and that they had neither suffered a seizure nor a head injury since their last appointment. At visits 5 and 7 an additional MMSE was performed.

Volumetric MRI scans were acquired using the protocol previously outlined. The same radiographer performed all scanning on the same 1.5T Unit. Each volumetric scan took less than 10 minutes. On the occasions when two scans were performed, the subjects were not removed from the scanner. Neither restraint nor sedation was used. If scans were clearly marred by excessive movement, at the discretion of the radiographer and with the consent of the subject and carer, a repeat scan was performed.
On completion of the seven scans, subjects were seen for a further visit (visit 8). At this time, repeat neuropsychology was performed, using the previously described battery; the diagnosis was reviewed; and assent for post-mortem examination was sought. Those subjects attending additional visits nine and ten were tested on the MMSE and underwent a further volumetric MRI scan.

Average annual percentage change in MMSE was calculated over the time points: 6 months, one year, 18 months and two years, for patients and controls. Estimated sample sizes required to power a study designed to detect a 20% reduction in MMSE with 90% power, and 5% significance (two-tailed) were calculated using standard formulae (see methods) (Fox et al, 2000).

Psychology scores were converted into z-scores and grade scores as previously described (see chapter one). Percentage changes in these scores (and the derived composite scores) were calculated. Power calculations were performed as above.

6.3 Results

6.3.1 Retention

The numbers of patients attending for each scan visit, and the number of scans acquired are shown in Table 6-2. Of the planned prospective imaging study (ie visits 1-7), 483 visits were planned. Of these, 463 (96%) were achieved to schedule. In the planned prospective imaging study (visits 1-7), 690 scans were scheduled. Of these, 647 (93%) were achieved.
Table 6-2. Study visit attendance

<table>
<thead>
<tr>
<th>Planned time (months)</th>
<th>Visit 1</th>
<th>Visit 2</th>
<th>Visit 3</th>
<th>Visit 4</th>
<th>Visit 5</th>
<th>Visit 6</th>
<th>Visit 7</th>
<th>Visit 9*</th>
<th>Visit 10*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numbers attending (patients)</td>
<td>69 (46)</td>
<td>66 (44)</td>
<td>67 (45)</td>
<td>68 (46)</td>
<td>66 (44)</td>
<td>60 (39)</td>
<td>67 (44)</td>
<td>39 (26)</td>
<td>22 (14)</td>
</tr>
<tr>
<td>Subjects scanned x1</td>
<td>3</td>
<td>66</td>
<td>4</td>
<td>68</td>
<td>66</td>
<td>3</td>
<td>67</td>
<td>39</td>
<td>22</td>
</tr>
<tr>
<td>Subjects scanned x2</td>
<td>65</td>
<td>-</td>
<td>63</td>
<td>-</td>
<td>-</td>
<td>57</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scans completed</td>
<td>133</td>
<td>66</td>
<td>130</td>
<td>68</td>
<td>66</td>
<td>117</td>
<td>67</td>
<td>39</td>
<td>22</td>
</tr>
</tbody>
</table>

*: Optional visits, not initially scheduled

During the course of the study, one patient dropped out, having been sectioned under the mental health act, and missed visits 2 and 5-7. One control suffered a heart attack, missing visits 4-6, but successfully returned for scan 7. The remaining missing 12 visits were due to unavoidable often health-related events; e.g. during the course of the study, 2 patients suffered broken limbs. On ten occasions, when two scans were scheduled, only one could be performed. Reasons included excessive movement on one scan, or late arrival at the scanner.

In the additional, extra phase of the study, 39 subjects (26 patients) attended for scanning at 18 months, and 22 subjects (14 patients) for scanning at one year. Thus in total, 524 subject visits were made, and 708 individual volumetric MRI scans were performed.
In order to remove one possible source of error, it was aimed to scan patients as close as possible to the scheduled intervals (see above). Table 6-3 demonstrates that in this study, this aim was achieved with a high degree of accuracy.

Table 6-3. Timing of achieved visits.

<table>
<thead>
<tr>
<th>Visit</th>
<th>Visit 2</th>
<th>Visit 3</th>
<th>Visit 4</th>
<th>Visit 5</th>
<th>Visit 6</th>
<th>Visit 7</th>
<th>Visit 9*</th>
<th>Visit 10*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planned interval from baseline (days)</td>
<td>-</td>
<td>14</td>
<td>42</td>
<td>84</td>
<td>182</td>
<td>266</td>
<td>365</td>
<td>546</td>
</tr>
<tr>
<td>Achieved mean (SD) interval from baseline (days)</td>
<td>-</td>
<td>16</td>
<td>43</td>
<td>98</td>
<td>180</td>
<td>270</td>
<td>365</td>
<td>552</td>
</tr>
</tbody>
</table>

*: optional visits, not initially scheduled

### 6.3.2 Clinical change

No subject suffered a seizure or significant head injury during the study. Although the majority of patients were stable either on or off cholinesterase inhibitors, the number on these drugs increased as the study progressed (Table 6-4).

Table 6-4. Acetylcholinesterase inhibitor use.

<table>
<thead>
<tr>
<th>Visit 0 &amp; 1</th>
<th>Visit 2</th>
<th>Visit 3</th>
<th>Visit 4</th>
<th>Visit 5</th>
<th>Visit 6</th>
<th>Visit 7</th>
<th>Visit 8</th>
<th>Visit 9*</th>
<th>Visit 10*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of patients on cholinesterase inhibitors</td>
<td>61%</td>
<td>64%</td>
<td>69%</td>
<td>74%</td>
<td>82%</td>
<td>87%</td>
<td>88%</td>
<td>88%</td>
<td>85%</td>
</tr>
</tbody>
</table>

*: optional visits, not initially scheduled
The mini-mental state examination was measured at visit 0, visit 5 (6 months), visit 7 (one year), and where available, visits 9 (18 months) and visit 10 (two years). Table 6-5 shows annualized change in MMSE for patients and controls for these intervals from the baseline assessment.

Table 6-5. Mean decline in MMSE score.

<table>
<thead>
<tr>
<th></th>
<th>6 months</th>
<th>12 months</th>
<th>18 months</th>
<th>24 months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients</strong></td>
<td>1.0 (5.9)</td>
<td>1.8 (3)</td>
<td>2.5 (2.5)</td>
<td>3.0 (2.6)</td>
</tr>
<tr>
<td>n=42</td>
<td>n=43</td>
<td>n=26</td>
<td>n=14</td>
<td></td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td>0.0 (1.8)</td>
<td>0.7 (1.1)</td>
<td>0.1 (0.7)</td>
<td>-0.1 (0.4)</td>
</tr>
<tr>
<td>n=21</td>
<td>n=22</td>
<td>n=13</td>
<td>n=8</td>
<td></td>
</tr>
</tbody>
</table>

Using these results, to power a study to detect a 20% reduction in MMSE with 90% power and 5% significance (paired), assuming no dropouts and not accounting for normal ageing, the following numbers of patients would be required in each arm of a placebo controlled trial: 18271 (6 months), 1458 (one year); 525 (18 months); 394 (two year).

When the patients were stratified on the basis of acetylcholinesterase inhibitor (AchEI) use (i.e. stable on, stable off, started AchEI during study or stopped AchEI), although there were only small differences between the baseline MMSEs, there appeared to be substantial differences in the MMSE change over the following year (with the caveat that the numbers in each group were relatively small) (Table 6-6). However, there was no significant difference in the MMSE change between those stable on AchEI and those starting treatment.
Table 6-6. Mean change in annual MMSE score and AchEI treatment.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD) Baseline MMSE</th>
<th>Mean (SD) annual change in MMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stable on AchEI throughout study</td>
<td>19.6 (4)</td>
<td>1.7 (3.1)</td>
</tr>
<tr>
<td></td>
<td>n=25</td>
<td>n=24</td>
</tr>
<tr>
<td>Stable off AchEI throughout study</td>
<td>16.7 (4)</td>
<td>5.6 (0.5)</td>
</tr>
<tr>
<td></td>
<td>n=3</td>
<td>n=3</td>
</tr>
<tr>
<td>Started AchEI during study</td>
<td>19.6 (4)</td>
<td>0.9 (3.6)</td>
</tr>
<tr>
<td></td>
<td>n=14</td>
<td>n=14</td>
</tr>
<tr>
<td>Stopped AchEI during study</td>
<td>16 (4.9)</td>
<td>2.5 (4.9)</td>
</tr>
<tr>
<td></td>
<td>n=2</td>
<td>n=2</td>
</tr>
</tbody>
</table>

6.3.3 Neuropsychological change

Serial neuropsychology was performed on 65 subjects (43 patients). The mean (SD) interval between the two assessments was 402 (± 30) days. Table 6-7 summarizes annualized mean (SD) change in each neuropsychological domain (positive results demonstrate an improvement; negative, a decline). Figure 6-1 demonstrates the baseline scores and changes in the various tests. Whilst the patients declined on all scores, the controls showed improvement in a number of tests, demonstrating a "practice effect" (Figure 6-1). Significant excess decline was seen in patients compared with controls for all tests apart from the recognition memory test for faces (p=0.5), and vocabulary scale of the WASI, which just failed to reach significance (p=0.07). Using the composite memory and cognitive scores as outcome measures, to power a study to detect a 20% reduction in the outcome measure with 90% power and 5% significance (paired), assuming no dropouts, 3732 patients would be required in each treatment arm using the composite memory scores, and 1013 using the composite cognitive scores.
Table 6-7. Annualized mean (SD) change in neuropsychological score

<table>
<thead>
<tr>
<th>Test Description</th>
<th>Controls</th>
<th>AD Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>RMT-words Grade score</td>
<td>22</td>
<td>0.25 (0.7)</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.08 (0.8)</td>
</tr>
<tr>
<td>Composite memory Grade</td>
<td>22</td>
<td>0.17 (0.5)</td>
</tr>
<tr>
<td>score change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silhouettes z-score</td>
<td>21</td>
<td>0.21 (0.5)</td>
</tr>
<tr>
<td>change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arithmetic z-score</td>
<td>21</td>
<td>0.03 (0.6)</td>
</tr>
<tr>
<td>change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naming z-score change</td>
<td>21</td>
<td>-0.11 (0.4)</td>
</tr>
<tr>
<td>Vocabulary z-score</td>
<td>22</td>
<td>-0.10 (0.4)</td>
</tr>
<tr>
<td>change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matrix reasoning z-score</td>
<td>22</td>
<td>0.07 (0.5)</td>
</tr>
<tr>
<td>change</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Composite cognitive</td>
<td>22</td>
<td>-0.01 (0.3)</td>
</tr>
<tr>
<td>z-score change</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*: significant difference, \( p<0.05 \), Wilcoxon signed rank test

†: significant difference, \( p<0.05 \), Student’s t-test
6.4 Discussion

6.4.1 Retention

The results of this study suggest that it is possible for patients with AD to undergo studies involving multiple short interval scanning. The patient drop-out rate was remarkably low (1/46), and the vast majority of scans were completed to schedule, at intervals very close to the study design.

Koss et al studied factors predictive of dropout in natural history studies of AD, and concluded that the level of site commitment was the most significant determinant of continued participation in a natural history study of AD; inclusion of both husband and wife where one is a patient and the other a control subject also increased retention (Koss et al, 1999). In this study both of these factors were addressed. Wherever possible, the patient’s spouse was recruited as a control, and a great deal of time was
spent establishing a personal rapport with each patient and carer. This included contacting each subject before each visit to provide both a reminder, and when appropriate, providing time for informal conversation. Other factors that may have improved patient retention in this study were: (1) the provision of realistic, comprehensive information as to exactly what the trial would entail, including giving the exact dates and time for all visits at the time of recruitment; (2) arranging for taxis to take patients to and from their scan; (3) providing a degree of flexibility in rearranging scans, should one be missed. It is also likely that the frequency of the visits over a relatively short time scale may have improved compliance, ensuring that subjects are kept regularly in touch with the research team. In longer studies, drop-out rates, even in healthy individuals are increased. Thus Rusinek reported drop out rates of 22% over two years in healthy subjects scanned twice in a prospective natural history imaging study of normal ageing (Rusinek et al, 2003).

It is of note that patients were not paid to enter the study, and received no new treatments. Mastwyk et al studied the reasons carers allowed patients with AD to enter clinical treatment trials, and found that hopes for a cure, contribution to medical science, making the affected relative feel better and keeping in regular contact with research staff were all reasons given (Mastwyk et al, 2002). There is also some evidence to suggest that taking part in trials may in itself be beneficial to the outcome in individual patients (Albert et al, 1997). In drug trials, patients and carers may have different expectations from those participating in purely research based trials. Whilst patients and carers may have an increased incentive to remain in a study if they feel that the treatment may be beneficial, this may be overshadowed by the effect of adverse drug reactions or side effects. However, the experience of this study and others suggests that close personal contact between investigator and patient and/or carer, combined with provision of appropriate information and planning may serve to increase retention in clinical trials in AD.

6.4.2 Clinical change

During the course of the study, there was an increase in the number of patients starting treatment with an acetylcholinesterase inhibitor. Whilst the aim at the start of the study was to attempt to recruit patients stable either on or off treatment, this was
not possible in all cases. During the course of the study, the availability of these drugs increased, and it followed that, as all patients in this study were eligible for treatment under guidelines issued by the UK National Institute of Clinical Excellence, increasing numbers of patients started these drugs.

The extreme variability of the MMSE changes over six months demonstrates the large variability inherent in the test, particularly over short intervals (Clark et al, 1999), as well as the associated ceiling and floor effects and non-linearity of the test (Galasko et al, 2000). Over periods of one year or more, the MMSE of the patient group fell at a rate of between 1.8 and 3.0 points per year, whilst in the controls this ranged from a gain of 0.1 to a loss of 0.7 points. In a recent study, Jack et al reported annual median decline in MMSE to be 1.9 points in AD (Jack et al, 2003). In a meta-analysis to assess the rate of change of MMSE in AD, Han et al calculated an average rate of change of 3.3 points, and in a cohort of 372 patients, Clark et al reported an annual decline of 3.4 points, with wide variability in rates (Clark et al, 1999). However, both these reports are likely to relate to subjects in the pre-acetylcholinesterase inhibitor era, and the slightly lower rates in this study, and in that by Jack et al, are likely to reflect the improved cognition produced by these drugs. Although the numbers are small, this appears to be reflected by the MMSE changes seen in the subgroup analysis. Those patients starting cholinesterase inhibitors had a trend for slower rates of decline than those stable on medication, or stopping treatment during the course of the study. Those not on cholinesterase inhibitors at all appeared to undergo decline at the fastest rate, with the caveat that patient numbers in this group (n=2) are very small.

The estimate that over 1400 patients would be required per arm to power a treatment trial based on the MMSE over a one year period, reflects that the MMSE was designed as a screening test for AD, and not as an accurate progression marker. This is further confirmed by Jack et al, who estimated sample sizes of 241 per treatment arm to detect a 50% treatment effect at a one-sided t-test at the 0.05 level (Jack et al, 2003). For a 20% treatment effect with a two-sided test (i.e. comparable to this study), this represents a sample size of 3036. Thus, whilst useful as a screening test, the MMSE is not a suitable sole progression marker for treatment trials in AD.
6.4.3 Neuropsychological change

These data clearly reflect the fact that AD is associated with inevitable cognitive decline, involving not only memory, but also other cognitive domains. This is reflected by the decline in both individual tests and the composite scores. A practice effect, evident for the majority of the tests in the controls was not seen in the patients, resulting in a further divergence of the already statistically significant scores at baseline. Notable exceptions to the pattern of significant decline compared to the controls are seen in the recognition memory test for faces and the vocabulary subset of the WASI. A reduction in scores on tests of memory for words but not faces has been shown to be predictive of progression from presymptomatic to symptomatic AD (Fox et al, 1998); this study suggests that there may similarly be a trend for slower progression in memory for faces than words in established sporadic AD. Moreover some patients with mild-moderate AD were already at floor on tests of memory. Conversely, the relatively small decreases in the vocabulary subset of the WASI observed in this study again reflect the relative preservation of this cognitive domain in the mild-to-moderate stages of AD.

These results demonstrate that it is possible to design a relatively short battery to assess several cognitive domains that may be applicable both to patients with mild-moderate AD and age-matched controls. Nonetheless, the estimated sample sizes that would be required to power a clinical trial are still too high to make the battery used in this study a viable trial outcome measures when used in isolation.
7 Scan quality

7.1 Introduction

If atrophy rates derived from serial MRI scans are to be used as outcome measures for clinical trials, the quality of the images acquired is critical. As previously discussed, volumetric MRI scans can be degraded for a number of reasons, which include those due to the subject (e.g. movement), those due to MRI acquisition (e.g. inadequate head coverage or wrap) and those due to imaging instrumentation (e.g. inhomogeneity).

In this study, a number of strategies were employed to minimize these potential sources of error. This chapter describes these strategies, and describes an assessment of the quality of the scans collected.

7.2 Methods

All scan were performed using the scanning protocol previously described, which was designed to provide scans with maximum signal-to-noise, high anatomical specificity, and minimum acquisition time. To further aid compliance and minimize movement, subjects were informed about the nature of the scanning, the requirement to keep as still as possible, and the noise associated with MRI scanner. Once in the scanner, soft foam pads were used to aid head stabilization, and foam ear pads were used, both for comfort and to diminish the possibility of cochlear damage (Radomskij et al, 2002). To minimize inadequate head coverage and wrap, the field of view was set to ensure adequate head coverage, and all patients were placed within the scanner by the same experienced radiographer. To reduce inhomogeneity, bias correction was performed on all scans using the N3 correction filter (Sled et al, 1998).

To assess the quality of the acquired scans, each volumetric scan was visually assessed and rated in two stages. In the first, movement artefact was assessed and graded as either: one (badly affected by movement); two (some movement); or three (no movement artefact). Examples of this grading are shown in Figure 7-1. Scans without movement artefact were then reassessed and subdivided three subgroups, (3, 4
or 5) based on a qualitative assessment of scan quality based on differentiation between white and grey matter, and brain and CSF (Table 7-1).

Table 7-1. Qualitative assessment of scan quality

<table>
<thead>
<tr>
<th>Grade</th>
<th>Movement</th>
<th>Tissue contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bad</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Present</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>None</td>
<td>Average</td>
</tr>
<tr>
<td>4</td>
<td>None</td>
<td>Good</td>
</tr>
<tr>
<td>5</td>
<td>None</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

The proportion of scans affected by movement artefact was compared between patients and controls. On visits with two back-to-back scans, the scan grades were compared between the first and second scans. Subsequently, each scan was re-reviewed to determine adequacy of head coverage and to exclude evidence of wrap or inhomogeneity. Head coverage was deemed inadequate if at least one posterior scan slice did not contain brain (Figure 7-2; a). Wrap made a scan inadequate if the misplacement of “nose” caused distortion at the “back” of the brain (Figure 7-2; b). A scan was rated as poor if inhomogeneity significantly altered brain:CSF differentiation across the scan.

Finally, when two scans were acquired on the same day, the rater determined which of the two was preferred; when both were graded the same, the first scan of the two was chosen.
Figure 7-1. Examples of scan grading

Grade 1: “Bad” movement artefact

Grade 2: “Some” movement artefact

Grade 3-5: “No” movement artefact
Figure 7–2. Examples of scans with inadequate head coverage and wrap.

(a) Inadequate head coverage

(b) Wrap
7.3 Results

Of the 708 individual volumetric MRI scans obtained, all could be analyzed (Table 7-2).

Table 7-2. Results of scan grading.

<table>
<thead>
<tr>
<th>Grade</th>
<th>Number of scans (percent)</th>
<th>Number of patient scans (percent)</th>
<th>Number of control scans (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16 (2)</td>
<td>13 (3)</td>
<td>3 (1)</td>
</tr>
<tr>
<td></td>
<td>56 (8)</td>
<td>44 (9)</td>
<td>12 (5)</td>
</tr>
<tr>
<td></td>
<td>115 (16)</td>
<td>90 (19)</td>
<td>25 (10)</td>
</tr>
<tr>
<td></td>
<td>243 (34)</td>
<td>163 (35)</td>
<td>80 (33)</td>
</tr>
<tr>
<td></td>
<td>278 (39)</td>
<td>155 (33)</td>
<td>123 (51)</td>
</tr>
</tbody>
</table>

The proportion of scans with bad movement (Grade 1) was ~ 3 times higher in the patients than the controls; however this was not a significant difference (Chi-square test, $p=0.19$). Controls were significantly more likely to have a scan without any movement (i.e. Grade 3-5) than the patients (chi-square test, $p=0.03$).

On the three time-points where two back-to-back scans were acquired (visits 1, 3, 6), the numbers of scans receiving each grade in the patient group is shown in Table 7-3. No significant differences in any scan grade were seen between the first and second scans.
Table 7-3. Grading of scans performed on the same day in AD.

<table>
<thead>
<tr>
<th>Grading</th>
<th>First scan</th>
<th>Second scan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade 1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Grade 2</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Grade 3</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>Grade 4</td>
<td>47</td>
<td>41</td>
</tr>
<tr>
<td>Grade 5</td>
<td>41</td>
<td>36</td>
</tr>
</tbody>
</table>

For non-movement related artefacts, 12 scans (2%) had inadequate head coverage, one had bad scan wrap, and one bad scan inhomogeneity. When scans with bad movement (Grade 1) were combined with those excluded on other grounds, a total of 30/708 scans (4.2%) were designated as “poor quality”.

7.4 Discussion

This study demonstrates that it is possible to obtain good quality serial volumetric MRI scans from patients with mild-moderate Alzheimer’s disease in a prospective trial setting. Whilst all scans obtained were of sufficient quality to be analysed, 95.8% were rated as being of good quality, increasing the chance of obtaining meaningful results from any future analyses.

Key to improving scan quality in a dementia trial is minimizing movement artefact, the most common cause of a poor scan in this study. A number of strategies are likely to have contributed to the small percentage of scans with significant movement artefact. In this study, the subjects taking part were highly motivated; the experienced radiographer played a crucial role not only in ensuring correct and reproducible positioning within the scanner, but also in reassuring patients when needed. Critically, as it is likely that the change of patient movement increases with the duration of the scan (Fox et al, 2000), the scanning protocol was designed to produce good quality scans in a time-frame tolerable to patients with AD (under 10 minutes).
It is notable that although overall controls were more likely than patients to have a scan with no movement artefact, patients had only a slightly (and not significantly) higher proportion (3 vs 1%) of “very bad” (Grade 1) scans due to movement than the controls. Furthermore, as the quality grading of the second scan was not significantly different from the first, these results suggest that should an individual move excessively on a first scan, it may be worth, with their permission, performing a second scan, which may be of considerably better quality.

Strategies to improve scan acquisition, such as ensuring an MRI protocol with a large enough field of view, using an experienced radiographer, and acquiring scans in a research MRI unit are all factors that are likely to account for the relatively few technical problems experienced in this project. The use of N3 inhomogeneity correction on all scans is likely to have reduced problems related to inhomogeneity: thus only one scan was deemed to be of poor quality because of inhomogeneity after N3 correction.

Few data exist on scan quality in longitudinal prospective AD studies. It is likely that in multi-centre drug studies, more problems will be encountered: subjects may be more variably motivated; radiographers may not be as aware of the needs of patients with dementia; protocols may not be strictly adhered to; software or hardware changes may occur during the course of the study; and differences within and between scanners may all increase the chances of poor quality scans. Nonetheless, many of the issues may be addressed at the trial design stage, and the use of a central image centre both to check image quality and request repeat imaging if necessary may improve the consistency of imaging between sites. This study demonstrates that, at least in a single centre, with sufficient planning and care it is possible to acquire good quality serial MRI scans in patients with AD.
8 A longitudinal case study of thalamic stroke

8.1 Introduction

Serial assessments of subjects allows for the demonstration of change over a period of time. Whilst this study was designed to assess progressive change in cohorts of patients with AD and normal controls, occasionally it may be possible to determine specific events or change occurring on an individual basis; this may be particularly the case when assessing relatively elderly patients over an extended period of follow-up. During the course of this study, one control subject suffered an acute dysphasic episode with persistent memory disturbance. By having imaging and neuropsychological data acquired before the onset of this new deficit, it was possible to demonstrate the changes that occurred subsequent to this event, and draw clinico-anatomical conclusions as a result.

8.2 Case report

8.2.1 Clinical assessments

ANC, a 68 year old right handed man volunteered and consented to take part as a control subject in this study. He was seen for his first assessment in February 2001, when he was well, and complained of no cognitive symptoms. He was a retired management consultant, whose past medical history included a hip replacement, mild renal impairment and benign prostatic hypertrophy, for which he was taking Finasteride. He did not smoke, and drank less than 21 units of alcohol per week. Examination revealed hypertension (BP 170/95) and a body mass index of 28kg/m², but was otherwise unremarkable. He scored 29/30 on the MMSE.

Baseline neuropsychology was performed (see chapter 8), and a volumetric MRI brain scan was acquired. Over the following nine months, he remained well and had five further volumetric MRI scans as part of the study. There was no evidence of sustained hypertension. He was commenced on Orlistat during this time by his general practitioner to help weight control. He scored 30/30 on the MMSE performed six months into the trial.
Ten months into the trial, he suffered an acute dysphasic episode, when he was suddenly unable to get the correct words out or to complete sentences. There was no associated limb or facial weakness. He retired to bed, and woke two hours later, by which time his speech had virtually returned to normal. Over the next few days, his wife reported him to be slightly confused, and to have continued difficulties in word finding, as well as new difficulties in remembering names. He saw his general practitioner who diagnosed a probable stroke, commenced treatment with aspirin, and referred him to the local neurology service.

One month later he attended for his final trial visit. He and his wife reported that his cognition had improved, but that although his ability to recognize faces was unimpaired, he had persistent difficulties recalling people's names. He had also become more reliant on his diary. On examination, the MMSE was 26/30, the blood pressure was 130/85, and a neurological examination was normal.

8.2.2 Imaging

All volumetric imaging was performed on the same scanner, using the previously outlined protocol. All scans were reported by an expert neuroradiologist. The first scan revealed evidence of mild small vessel disease especially affecting the pallidum, with normal hippocampi and no evidence of global or regional cerebral atrophy. The next five scans (over the following nine months) revealed neither significant atrophy nor any new ischaemic lesions. The final scan, one month after the acute event, revealed a discrete new infarct in the left thalamus, involving the medial thalamic nuclei, and interrupting the mamillo-thalamic tract (Figure 8-1).
Figure 8–1. Demonstration of new thalamic lesion.
Baseline (a) and one-year scan (b) are shown. Lesion is highlighted (arrows).

8.2.3 Neuropsychological assessments

ANC was assessed at the start and end of the study using the standard neuropsychological battery, as previously described. A summary of these results is shown in Table 8-1. At the start of the study, he performed in the superior range on tests of verbal and non-verbal intelligence, recognition memory, reading and picture naming. Calculation was in the average range, and visuoperceptual were satisfactory. Reassessment was carried out at one year, a month after the acute event. The only change noted was a subtle decline in naming skills on the Graded Naming Test (McKenna and Warrington, 1983); nonetheless, his score still fell within the superior range (90th percentile). In light of his memory complaint, additional tests of word retrieval skills and episodic memory were administered.
Table 8-1. Baseline and one-year neuropsychology scores for ANC.
Raw scores and percentile rankings are shown.

<table>
<thead>
<tr>
<th>Neuropsychological tests</th>
<th>Assessment 1 (Feb 2001)</th>
<th>Assessment 2 (Feb 2002)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WASI vocabulary (raw score)</td>
<td>77/80</td>
<td>76/80</td>
</tr>
<tr>
<td>WASI matrix reasoning (raw score)</td>
<td>28/32</td>
<td>29/32</td>
</tr>
<tr>
<td>Predicted full scale IQ</td>
<td>134 (99%ile)</td>
<td>135 (99%ile)</td>
</tr>
<tr>
<td>NART Predicted IQ</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>RMT-Words</td>
<td>25/25 (&gt;50%ile)</td>
<td>25/25 (&gt;50%ile)</td>
</tr>
<tr>
<td>RMT-Faces</td>
<td>25/25 (&gt;50%ile)</td>
<td>25/25 (&gt;50%ile)</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>14/24 (&gt;50%ile)</td>
<td>14/24 (&gt;50%ile)</td>
</tr>
<tr>
<td>Naming</td>
<td>29/30 (&gt;99%ile)</td>
<td>25/30 (&gt;90%ile)</td>
</tr>
<tr>
<td>Silhouettes</td>
<td>21/30 (&gt;25%ile)</td>
<td>21/30 (&gt;25%ile)</td>
</tr>
</tbody>
</table>

Language skills: Two further stringent graded difficulty naming tests were attempted. ANC was able to name 22/30 objects and 28/30 animals (>50th percentile and >90th percentile, respectively) from the McKenna Category Specific Names test (McKenna, 1997). On a comparable proper noun retrieval test (historical figures, countries and buildings) (McKenna and Warrington, 1980) his performance was extremely competent (25/30). His good verbal comprehension was demonstrated on a test of knowledge of synonyms (Warrington et al, 1998) (concrete words: 22/25, >50th percentile; abstract words: 25/25, >90% percentile). ANC also expressed himself fluently using a wide vocabulary. Thus at this stage there was no evidence of a dysphasic syndrome.
Episodic memory: The routine neuropsychological battery contained only the Easy Recognition Memory Test on which ANC scored at ceiling. He was also tested on the standard Recognition Memory Test (Warrington, 1984), which allows visual and verbal memory to be assessed independently with an identical test design for each component. He scored 41/50 on the visual section of the test (Faces: 50th percentile). By contrast, on the verbal section of the test he scored only 36/50 (Words: 5th percentile). This discrepancy of five points represents a selective verbal memory deficit (p<0.02) (Warrington, 1984). On a demanding test of visual recognition memory, the Topographical Recognition Memory Test (Warrington, 1996), he scored 26/30 (75th percentile).

ANC’s performance was also impaired on a measure of verbal recall; he scored at the 5th percentile on the Camden Paired Associate Learning Test (CPALT) (Warrington, 1996). In an attempt to document this recall deficit in more detail, ANC was assessed with part of a longer famous faces test previously used by Cipolotti et al (Cipolotti et al, 2001) to provide evidence of a retrograde amnesia in their patient VC. This test consists of 39 monochrome photographs of famous people from the 1990s including politicians, entertainers, sportsmen and individuals famous because of particular newsworthy events. The subject was first requested to recall the name of each individual orally. Subsequently, each photograph was re-presented with a choice of three names in a forced choice recognition paradigm - the target name alongside two equally famous distractor personalities. His scores were compared with those of 20 age-matched control subjects tested by Cipolotti et al (Cipolotti et al, 2001). He was able to recall only 21% of names correctly, significantly fewer than the control subjects (50%; Z = 2.0, P < 0.02, 2-tailed test). However, no such difference was found using the forced choice measure (ANC 92%, controls 85%). These findings corroborate the evidence from the anterograde memory tests, providing further evidence for a selective verbal memory deficit.
8.3 Discussion

In this individual (ANC), using longitudinal imaging, it was possible to demonstrate the appearance of a new, discrete left sided thalamic infarct. At follow-up, both ANC and his wife reported memory deficits occurring during the period we documented this lesion to have appeared. Although he scored at ceiling on the Easy Recognition Memory Test at both the start and end of the study, his new reported memory impairment prompted us to undertake more detailed testing at follow-up. Clear, focal cognitive deficits were determined at this assessment, which are highly likely to be consequent to the localized thalamic infarct.

Thalamic lesions may produce a wide range of neuropsychological deficits. The memory dysfunction associated with thalamic lesions appears to be best correlated with disruption of the mamillo-thalamic tract, as seen in this case; the role of the medial thalamic nuclei in memory dysfunction is less clear (for a review, see Van Der Werf YD et al (Van der Werf et al, 2000). Subsequent to his infarct ANC developed a selective episodic memory impairment for verbal material with preservation of visual memory, as demonstrated by his performances on the Recognition Memory Test, Camden Paired Associate Learning Test and Topographical Recognition Memory Test, and his poor recall of names of famous faces despite good recognition. There was no evidence of semantic memory impairment as evidenced by the high score on the Graded Naming Test. Thalamic lateralization (where dominant lesions result in verbal memory impairment, and non-dominant in visual memory problems), has been proposed by several investigators (Clarke et al, 1994; Parkin et al, 1994; Mori et al, 1986; Parkin et al, 1994; Mori et al, 1986), although other authors have failed to demonstrate a consistent effect (Exner et al, 2001; Wallesch et al, 1983; Wallesch et al, 1983). These findings support the hypothesis that lateralization of cognitive processing of visual and verbal material exists at the thalamic as well as cortical level. Whilst lesions throughout the dominant hemisphere can cause a selective impairment of verbal memory (Warrington, 1984), it is rare for such lesions to be demonstrated longitudinally.
This case demonstrates that small thalamic lesions can cause persistent lateralized memory dysfunction in the absence of cortical signs or abnormalities on conventional imaging. Small lesions such as this could easily be missed on conventional imaging, and suggests that higher resolution scanning afforded by volumetric MRI may be a useful diagnostic tool in such circumstances. More generally, by using the subject as his own control, this case demonstrates the power of serial measurement in the detection of change and its correlates.
9 Stability in longitudinal scanning

9.1 Introduction

The use of atrophy rates derived from serial volumetric MR imaging as outcome measures in trials of novel treatments in AD is dependent on good quality serial scan acquisition; the results from chapter 7 suggest that this is possible in serial MRI studies in AD. Given that the magnitude of changes that might be expected in a treatment trial is small, it is also critical that MRI acquisition is stable throughout the study. One common cause of variation in image acquisition is ‘voxel dimension drift’, an unpredictable change in voxel size over time. Such changes are generally not visible on a one-off scan, but can have profound implications on measures of atrophy when two scans are compared. Changes in voxel size in an individual scanner can occur over time for a number of reasons, including scanner miscalibration or fluctuation in the performance of scanner gradients. These ‘dimension drifts’ in voxel size can result in up to 3% changes in apparent volume (Freeborough et al, 1996), similar to the rates of annual brain loss seen in AD. Errors introduced in this way may profoundly affect the power of cerebral atrophy as an outcome measure in clinical trials, particularly in studies involving short interval scanning.

Several methods have been developed to counter changes in voxel size over time. Test objects (phantoms) can be scanned serially, and relative scanner fluctuations detected either from measurements of the dimensions of the phantom, or from scaling factors provided by registering the serially acquired phantom images (Lemieux and Barker, 1998). However, the accuracy of correction depends on the precision to which the test object can be measured, and unless a phantom is scanned concurrently with each acquisition it is not possible to adjust for drift in voxel size on a scan-by-scan basis. Total intracranial volume (TIV), an estimate of premorbid brain volume which is unaffected by neurodegeneration (Jenkins et al, 2000), can be calculated on every scan acquired, and can be used to help correct for voxel size drift as the skull provides an invariant structure which acts as a natural ‘test object’. Skull or TIV-based correction for voxel size variation has been shown to improve the consistency of longitudinal measures (Whitwell et al, 2001) and increase the sensitivity of registration to detect brain atrophy (Freeborough et al, 1996). TIV measures are,
however, only semi-automated and are operator intensive, particularly in a study of this magnitude. In multi-centre clinical trials it is clearly advantageous to use a technique that is automated, fast, and as unbiased as possible.

An alternative means of correcting for voxel drifts, is the use of a 9-degrees-of-freedom (9dof) registration (see 2.3.1.1). The 9dof registration process combines a 6dof rigid registration (3 translations and 3 rotations), which matches the position and orientation of two brain volumes acquired from serial scans to a sub-voxel degree of accuracy (Freeborough et al, 1996), with an additional 3 scaling parameters (spatial scaling factors in three dimensions: x,y,z) to correct differences in voxel size (Woods et al, 1992; Woods et al, 1993). In theory, this process should compensate for variations in voxel size between pairs of scans without “scaling away” real atrophy, on the basis that the brain’s inherently complex shape (including its gyral foldings and central ventricles), atrophies in a way that cannot be corrected for by a linear scaling alone.

This work describes a critical evaluation of the relative merits of 9dof registration and TIV normalisation in correcting for longitudinal fluctuations in voxel size. In the first set of experiments, scans from control and patients with AD were artificially scaled, to:

1. Assess the ability of the 9dof registration to recover scaling changes;
2. Investigate the hypothesis that the scaling correction in the 9dof registration algorithm does not alter atrophy rates obtained from an automated atrophy quantification; and
3. Compare the accuracy of the 9dof registration scaling correction compared to normalisation using TIV.

TIV correction and 9dof registration were then assessed in a subset of the study cohort, (including both patients with AD and controls) scanned longitudinally over a variety of intervals ranging from two weeks to six months. During the course of this clinical study, two transient systematic changes in voxel size that had otherwise gone unnoticed were discovered; the timing and causes of these problems could be
accurately established, and the ability of both these techniques to overcome them could be demonstrated.

9.2 Methods

9.2.1 Subjects

Eighty-one scans from 23 subjects (11 controls and 12 patients with AD) taking part in the larger study were included. A variety of scan numbers and intervals were assessed: eight subjects with two scans; six with three scans; seven with four scans; and two with five serial scans. The scan intervals ranged from 10 to 182 days. The one year baseline scans from nine of the patients with AD were also assessed.

9.2.2 MRI acquisition, Phantom scanning and data analysis

All scans were performed on the same scanner, using the parameters and protocols previously described. The scanner quality assurance (QA) consisted of a structured programme in which all quantitative measurements were monitored by imaging suitable phantoms. This program was supplemented by regular imaging of normal control subjects. To assess geometric distortion, the QA programme consisted of scanning a cylindrical phantom (test object 2) from the Eurospin II Magnetic Resonance Quality Assurance Test Objects kit (Diagnostic Sonar Ltd, UK) using a scan sequence comparable to that used in this study. Geometric distortion is determined by transferring the image data to a Sun workstation (Sun Microsystems, Mountain View, CA, USA) and comparing the test object dimensions to a “reference” image. This is facilitated by the use of a locally written image registration programme, MReg. The post processing and analysis has been previously described (Lemieux and Barker, 1998). The degree of geometric distortion in the x, y and z axis, as compared with the reference scan, is expressed as a correction factor which would be necessary to normalise the image dimensions.

9.2.3 Image processing

All measurements were performed using the MIDAS image analysis software package (Freeborough et al, 1997). Whole brain volumes were delineated as previously
described (Freeborough et al, 1997). Serial scans were positionally matched to the baseline image using a 9dof registration (Freeborough et al, 1996; Woods et al, 1992; Woods et al, 1993). The scaling parameters within this algorithm define the voxel 'stretch' necessary to match image sizes, thereby accounting for any differences in voxel sizes. TIV was measured using a previously described and validated technique (Whitwell et al, 2001). Each unregistered image was put into the orientation defined by the Montreal Neurological Institute 305 brain average (Mazziotta et al, 1995) using a 6dof registration in order to improve the reproducibility of the TIV technique. Every 10th axial slice was segmented with the inferior border set as the lowest section in which cerebellar tissue was present. Linear interpolation of areas was used to obtain an estimate of the TIV from the segmented sections.

9.2.4 Recovering artificial scaling changes

The baseline scans, along with the outlined brain regions, for ten of the control subjects were artificially scaled (stretched) in the x, y and z dimensions. The same scalings were applied in each dimension: 0.5, 1.0, 1.5 and 2%. These artificial scalings resulted in a total volume change of 1.5, 3.0, 4.6 and 6.1% per scan respectively. A 9dof registration was used to register the scaled brains back onto the original baseline brain. The volume scaling factor (the product of the 3 linear scaling factors in 3 different dimensions, xyz) for each of these intervals was calculated from the 9dof registration.

9.2.5 Atrophy quantification in AD with and without scaling change

The nine, one-year scans (including the outlined brain regions), from the patients with AD, were artificially scaled by 1.0% in each of the x, y and z dimensions, resulting in a total volume change of 3.0%. The resultant 'scaled' one year scans were then registered to the relevant baseline scan. Global brain atrophy was calculated from the registered scans using the brain boundary shift integral (BBSI), as previously outlined (Freeborough and Fox, 1997). Atrophy rates derived from the original and scaled one-year to baseline scans were compared.
9.2.6 Comparison of TIV and registration using artificially scaled scans

As in 9.2.4, the baseline scans, along with the brain regions, for ten of the control subjects were artificially scaled in the x, y and z dimensions by 1.5%, resulting in a volume change of 4.6%. TIV was measured on the baseline scans and the artificially scaled scans. The TIV ratio (ratio of the TIV of each serial scan to baseline TIV) was calculated and compared to the volume scaling factor generated from the 9dof registration.

9.2.7 Comparison of TIV and registration using patient and control scans

Serial scans (excluding the one year scans i.e. a total of 49 scan pairs) for each of the 23 subjects were compared to baseline. TIV was measured on all scans. Each serial scan was registered (9dof) to the baseline scan.

9.2.8 Statistical analysis

The TIV ratios and volume scaling factors were compared using the Wilcoxon signed rank test. Pitman’s test was applied to test for differences in standard deviation between groups. The scaling factors generated from the serial scans from AD and control subjects clearly separated into three distinct groups. In order to compare the variability of the TIV and scaling factors for the entire group, it was necessary to normalize the values from the three groups. Thus, for each of the three groups, the standard deviations of the scaling factors and TIV ratios were divided by the relevant group-specific mean. Pitman’s test was then performed on this variable with all data points included in the analysis.

9.3 Results

9.3.1 Recovering artificial scaling changes

Figure 9-1 demonstrates that volume scaling factors generated from the 9dof registration were highly correlated to the known (artificial) volume scaling change (R=1.00, p<0.001). There was almost complete recovery of the imposed artificial changes; the mean difference between the volume scaling factors (xyz) from the registration and the known induced scaling changes (xyz) was very small: at an
artificial volume change of 1.5%, the mean obtained volume change was 1.4%, mean squared error (MSE) 0.016; at 3.0%, the mean obtained was 3.1%, MSE 0.016; at 4.6%, the mean obtained was 4.6%, MSE 0.004; and at 6.1% the mean obtained was 6.2%, MSE 0.007.

Figure 9–1. Artificial scaling factor plotted against 9dof recovered scaling factor.
The baseline scan from each subject was artificially scaled by 1.005 (0.5%), 1.01 (1%), 1.015 (1.5%) and 1.02 (2%) in each dimension (x, y and z) leading to an expected volume scaling factor (xyz) of 1.015, 1.030, 1.046 and 1.061 respectively. The resultant scans were registered to the original baseline scan.
9.3.2 Atrophy quantification in AD with and without scaling change

The mean loss of brain volume over a one-year interval in nine of the AD patients was 20.4 ± 9.6ml. Artificially scaling the repeat scans by 3.0% and then using a 9dof registration to correct for this change resulted in a very similar calculated mean volume loss of 19.4 ± 9.7ml.

9.3.3 Comparison of TIV and registration using artificially scaled scans

At an artificial scaling factor of 4.6% (1.5% in each dimension) there was no significant difference between the 9dof registration derived volume scaling factors (1.0461 ± 0.0004) and TIV ratio (1.0433 ± 0.0042; p=0.06). However the variance of the 9dof-derived volume scaling factors was significantly less than that of the TIV ratio (p<0.001, Figure 9-2).

Figure 9-2. Artificial scaling change correction using TIV and 9dof.
The volume scaling factors for both TIV correction (TIV2/TIV1) and 9dof registration (xyz) on ten control scans after an artificial scaling change of 4.6% (1.5% in each dimension) are shown. The known (artificial) volume scaling factor is indicated by a dashed line.

![Graph showing volume scaling factors for TIV and 9dof registration](image-url)
9.3.4 Comparison of TIV and registration using patient and control scans

The TIV ratio and 9dof derived volume scaling factors were highly correlated (R=0.98, p<0.001) with a mean difference of -0.0009. When the variability of the two methods was compared, the estimated standard deviation of the scaling factors was significantly less (by approximately one third) than that of the TIV ratios (serial scans to baseline TIV, i.e. TIV2/TIV1) (p<0.001).

The total voxel size drifts, determined by both the volume scaling factors and TIV ratios, were not normally distributed as had been expected, but surprisingly fell into three distinct groups (Figure 9-3): those with volume change of approximately 1 (>0.99, <1.01, n=34), and two smaller groups with volume changes around 0.97 (>0.96, <0.98, n=13) or 1.03 (>1.02, <1.04 n=2).

When the volume scaling factors were plotted against the period over which the scans were acquired (Figure 9-4) it was possible to identify two time points at which systematic changes had occurred. The first (labelled A on Figure 9-4) occurred between the 23rd and 27th of September 2000, and resulted in scaling increases of approximately 3%, implying a concurrent increase in voxel size by this amount. The second (labelled B) occurred between the 26th of October and 1st of November 2000 and resulted in a decrease of scaling factors (and therefore voxel size) by approximately 3%, i.e. correcting the original shift. It is notable that all scans encompassing only one of these time points have scaling factors of approximately +3% or -3%, whereas all scan pairs encompassing neither or both of these times have scaling factors approaching unity. The phantom quality assessment (QA) from the scanner is also shown in Figure 9-4, and demonstrates the same fluctuations in voxel size implied by these results. Significant shifts in voxel size occurred only in the x and y dimensions; the z (antero-posterior) dimension remained unchanged. The x (± 1.5%) and y (± 1.5%) dimension changes account for the volume changes of ± 3%. These results mirror the time course and magnitude of the changes suggested by the TIV or scaling changes.
Figure 9-3. 9dof registration volume scaling factor versus TIV ratio.
Figure 9–4. Correction factors from the scanner QA and 9dof registration.
The QA correction factors (upper panel) are plotted for the x (x), y (●) and z
dimensions (○) and represent the degree (%) of geometric distortion required to
normalise the image dimensions to a reference scan. The 9dof registration derived
volume scaling factors (lower panel) are given as a percentage. Each scan pair (n=49)
is associated with a scaling; both time points (baseline and repeat) are assigned this
value and plotted with a line joining them. Two time points at which systematic
changes in voxel size had occurred are illustrated: the first (A) occurred between the
23rd and 27th September; the second (B) between the 26th October and 1st November.
These changes are also seen in the plot of correction factors, with both the x and y
dimensions showing a 1.5% increase in dimension between September 2000 and the
middle of October 2000 and a consequent decrease at the end of October. The z
dimension remained relatively unchanged.
9.4 Discussion

This work clearly demonstrates that scanner related changes may be of the same order of magnitude as volume change due to neurodegeneration over a one-year period, and if uncorrected can significantly alter volume measurements made from serial MRI scans. These results demonstrate that scaling factors automatically generated from a 9dof registration can correct for these fluctuations in voxel size as well, or better, than TIV measurements in longitudinal MRI studies.

The experiments using artificially scaled voxel sizes showed that the 9dof registration can recover substantial scaling changes (ranging from 1.5 to 6.1%), whilst preserving the true atrophy occurring between two serial scans. Previous studies have established that measures of TIV are unaffected by brain atrophy (Jenkins et al, 2000; Whitwell et al, 2001); this is unsurprising given that the measure is based on the invariant skull boundary. By contrast the use of a 9dof registration to correct for voxel size drift might theoretically affect measures of atrophy; in particular 9dof registration might “remove atrophy”. These results suggest that, as with measures of TIV, the scaling factors generated by 9dof registration are not influenced by ongoing atrophy in AD. This is likely to be because brain atrophy in AD involves complex, non-linear changes with ventricles enlarging outwards and individual sulci selectively enlarging – changes that cannot be matched by the application of simple linear scalings (Freeborough and Fox, 1998; Scahill et al, 2002).

The 9dof registration corrects for drifts in voxel size at least as well as TIV correction. The smaller variance in the registration scaling factors in both the artificial and ‘real’ situation provides evidence to suggest that the 9dof registration may in fact be the more precise of the two methods. This may be due to the degree of operator error inherent in a semi-automated TIV measurement; TIV may also be affected by inadequate head coverage or changes in chemical shift. Furthermore, TIV measurement is time consuming, taking on average 20 minutes per scan; thus in the study in its entirety, this would equate to an extras 236 hours analysis time. Whilst accuracy of TIV correction might be improved by performing more measurements per scan (Eritaia et al, 2000), this would come at increased cost in terms of operator time. By contrast, 9dof registration is an automated technique, not significantly increasing
post-processing time, less reliant on segmentation quality, affected neither by operator bias nor error, and utilizing the entire brain volume.

As a consequence of acquiring multiple scans at different, short intervals, it was possible to identify accurately the timing of two systematic changes resulting in voxel size shifts of approximately 3%. As the annual rate of whole brain atrophy in AD is of the order of 2-3% (Fox and Freeborough, 1997), it is clear that such changes, if uncorrected, would seriously compromise the use of serial scanning to determine volume change (and consequently atrophy of whole brain or regional measures such as the hippocampus) as outcome measures in clinical trials of neurodegenerative diseases. The use of either 9dof registration or correction for TIV was successfully able to correct for these changes. By consulting the scanner log, it was possible to determine that the initial increase in voxel size occurred following a routine service on the 27th September 2000 by an engineer who did not regularly work on this scanner. The subsequent reduction of voxel size (i.e. correction to baseline) occurred at the next routine service on the 27th October 2000, this time carried out by the regular engineer. Whilst neither engineer noted any specific scanner problems, the scanner radiographers identified a fluctuation in QA after the first service and informed the regular engineer before the second service. It is important to note that this problem is unlikely to be unique to this particular scanner; drifts in voxel size, and the potential for miscalibration are generic problems affecting all MRI scanners. Although this shift in voxel size was demonstrated in the weekly QA phantom scan, it would not have been possible to correct each individual scan accurately on this basis. Either the scans would have to have been discarded or the patient recalled. As MRI-based measures of atrophy are increasingly used in multi-centre trials, there are clear advantages in using a post-processing method to allow retrospective correction both for drift in voxel size and for the possibility of scanner miscalibration, thus allowing a higher proportion of acquired scans to be included in the analysis. This does not however negate the need for regular quality assurance and extreme care in scan acquisition.

Stability of MR acquisition is fundamental to the accurate measurement of cerebral atrophy rates. In trials in which brain atrophy is an outcome measure, the potential for acquisition instability needs to be recognized and minimized. Routine scanner
services may introduce significant acquisition changes and whilst quality assurance is an essential component of all serial MRI studies (Leary et al, 1999) post-acquisition correction methods are likely to be critical to the success of MR-based outcome measures in multi-centre studies. 9dof registration can correct for voxel drifts without altering atrophy measures, and is unbiased, automated, quicker and less variable than correction factors based on manual measures of TIV. As either method was shown to correct for voxel drift without affecting measures of atrophy, the automated 9dof, with the associated saving in image analysis time, was used to correct for scanner drift in the larger study.
10 Ventricular volumes and atrophy

10.1 Introduction

Global brain atrophy, as previously discussed, is an inevitable feature of the progression of AD. The results of brain atrophy may be visible on a one-off MRI scan as cortical thinning, and as ventricular enlargement. However, inter-individual variation in head and ventricular size ensures that there is likely to be substantial overlap between patients and controls in cross-sectional studies (Bradley et al, 2002; Whitwell et al, 2001). Serial MRI acquisition and measurements can quantify the amount of volume loss occurring between scans, and potentially provide a signal that can be used in a clinical trial; a truly disease-modifying (rather than symptomatic) agent would be expected to slow excess atrophy in the treated group.

A number of ways of measuring such atrophy have been proposed. These include manual measures (that will be considered in this chapter), as well as (semi-)automated methods, such as the BBSI (Freeborough and Fox, 1997), which will be considered in later chapters. Few studies have assessed global brain atrophy by any of these means at intervals of under one year, and fewer still have done so in a prospective manner that might approximate to a clinical trial.

Since the early days of CT, ventricular size and rate of change have been used to discriminate Alzheimer’s disease from normal ageing (Luxenberg et al, 1987; de Leon et al, 1989). These early studies highlighted the overlap in cross-sectional ventricular volumes between patients and controls, and concluded that rate of change was a better discriminator. Several later studies using MRI confirmed these original findings, both cross-sectionally (Murphy et al, 1993) and longitudinally (Silbert et al, 2003; Jack et al, 1998; Wang et al, 2002). These studies were often performed over a number of years, and with the exception of Bradley et al, who reported results of a combined ventricle-to-brain ratio (Bradley et al, 2002), few studies have reported ventricular volume changes over intervals under one year. The aim of this work was to compare cross-sectional and longitudinal ventricular measures as diagnostic markers; to determine ventricular atrophy rates over one year and six months as potential outcome
measures; and to compare these ventricular volume and volume change with ratios of ventricle-to-brain volume.

10.2 Methods

Of the previously described cohort, scans from 38 patients with AD and 19 controls were used for this study. Scans at three time points were used: baseline (visit 1), 6 months (visit 5), and one year (visit 7). For the baseline scan, the best rated of the two scans was used. Subjects were included if scans from these intervals were available for analysis.

Scan acquisition and pre-processing

Scans were acquired as previously described, and transferred to Sun workstations for analysis. Scans were segmented to exclude non-brain tissue using the previously described protocol (Freeborough et al, 1997). Once all brain regions were highlighted, for consistency and to improve the reproducibility of the ventricular measurements, all scans were put into the orientation defined by the Montreal Neurological Institute 305 brain average (Mazziotta et al, 1995) using a 6dof registration. To avoid problems associated with voxel drift (see chapter 9), the six month and one year scans thus derived were both registered to the baseline registered image, this time using a 9dof registration (Freeborough et al, 1996).

Ventricular segmentation

All ventricular volumes were measured using the MIDAS software package (Freeborough et al, 1997). Volumes were measured on the baseline and two repeat images for each subject using 60% of mean brain image intensity as an upper threshold. Ventricular volumes included the lateral ventricles and temporal horn of the lateral ventricles but not the third or fourth ventricle. All segmentations were performed by one of two raters (Shona Price or myself); mean inter-rater and intra-rater reproducibility was >99.5%. Further details of the ventricular segmentation protocol are given in Appendix 7.
Statistical analysis

Ventricular volumes at baseline were compared between patient and controls using non-parametric statistics, as they were not normally distributed (Skewness and Kurtosis test statistic < 0.05). Rates of ventricular expansion were calculated by simple subtraction of volume at baseline from volume at follow-up, corrected for inter-scan interval, and expressed as ml/year. All longitudinal results were normally distributed (not significantly skewed), and therefore comparisons were made using t-tests. Associations between age and volumes were performed using linear regression. The ratio of ventricle-to-brain volume was calculated at each time point and expressed as a percentage (i.e. 100*(ventricular volume/brain volume)); this ratio was compared between patients and controls at baseline, and the rate of change of this index (expressed as a percentage/annum) was used to assess progression over both one year and six months. Sample size estimates of ventricular change and ventricle-to-brain ratio for each interval were calculated to detect a 20% reduction in atrophy with 90% power and 5% significance (paired), assuming no dropouts. In addition, correction was made to account for the effect of normal ageing (Fox et al, 2000).

10.3 Results

10.3.1 Ventricular volumes

10.3.1.1 Cohort demographics

The patient cohort consisted of 23 women and 15 men; in the controls there were 10 women and 10 men. The groups were well matched for age: patients mean (SD) 69.8 (7) years; controls 69.1(7) years.

10.3.1.2 Cross-sectional

Ventricular volumes were significantly larger in patients and controls (p=0.0001, Wilcoxon rank sum test) (see Table 10-1). However, there was substantial overlap between the groups (Figure 10-1). Ventricular volume could distinguish patients from controls with a sensitivity of 81%, and a specificity of 68%.
Table 10-1. Baseline ventricular volumes.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean (median) volume /ml</th>
<th>Range / ml</th>
<th>Interquartile range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>38</td>
<td>52.7 (46.0)</td>
<td>21.4 – 129.1</td>
<td>29.8</td>
</tr>
<tr>
<td>Controls</td>
<td>19</td>
<td>31.7 (24.9)</td>
<td>15.2 – 113.4</td>
<td>18.2</td>
</tr>
</tbody>
</table>

Figure 10-1. Baseline ventricular volumes for patients and control (ml)

There was a significant relationship between baseline age and baseline ventricular volume for both patients ($p=0.0039$) and controls ($p=0.046$).
10.3.1.3 Longitudinal change

Change over one year

The mean (SD) length of follow-up between scans was 365 (14) days. The rate of ventricular change was significantly greater in patients than controls ($p<0.0001$) (Table 10-2; Figure 10-2). The sensitivity and specificity of ventricular volume change to classify patients from controls were 85 and 95% respectively. There was no significant increase in rate of change with increasing age for either group.

Table 10-2. Annual change in ventricular volume (ml) over one year.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean ml/ Year (SD)</th>
<th>Range</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>38</td>
<td>4.95 (2.5)</td>
<td>1.2 – 10.3</td>
<td>4.1 – 5.8</td>
</tr>
<tr>
<td>Controls</td>
<td>19</td>
<td>1.04 (1.05)</td>
<td>-0.8 – 3.2</td>
<td>0.5 – 1.6</td>
</tr>
</tbody>
</table>

Figure 10-2. One year ventricular change for patients and controls.
Change over six months.

Over both six month periods, there was significant excess ventricular enlargement in the patients compared to the controls \((p<0.0001)\), and both the mean and standard deviation of the annualized rate of ventricular increase were very similar to those seen over one-year (Table 10-3). The sensitivity was 87% and specificity 75% for the first period, and 86% and 68% for the second.

Table 10-3. Ventricular change over the two six month periods.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Mean (SD) interval (days)</th>
<th>n</th>
<th>Mean ml/year (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6 months</td>
<td>180 (7)</td>
<td>38</td>
<td>4.22 (2.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>0.78 (1.3)</td>
</tr>
<tr>
<td>6-12 months</td>
<td>186 (14)</td>
<td>38</td>
<td>5.67 (3.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>1.29 (1.2)</td>
</tr>
</tbody>
</table>

Power calculations

Based on the assumptions previously outlined, (power: 90%; sensitivity, 5% level – paired; placebo controlled trial), assuming no drop outs and no unusable scans, at one year 137 patients would be required in each treatment arm. Based on the results of the first six month time period, this number would be 170 per arm; on those from the second: 200. Accounting for normal ageing, these results were 220 over one year, and 255 and 335 per arm over 6 months.

10.3.2 Ventricle-to-brain ratio

10.3.2.1 Cross-sectional

Ventricular-to-brain ratios were significantly larger in patients than controls \((p<0.001,\) Wilcoxon Rank Sum test) (Table 10-4). Compared to ventricular volume alone, the sensitivity was slightly improved (87%, specificity of 68%).
Table 10-4. Baseline ventricular-to-brain ratio
Where ventricular-to-brain ratio = 100* (ventricle volume/brain volume)

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean (Median)</th>
<th>Range</th>
<th>Interquartile range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>38</td>
<td>5.5 (5.0)</td>
<td>2.2 - 17.1</td>
<td>3.15</td>
</tr>
<tr>
<td>Controls</td>
<td>19</td>
<td>2.9 (2.3)</td>
<td>1.3 - 10.1</td>
<td>1.42</td>
</tr>
</tbody>
</table>

10.3.2.2 Longitudinal change

Change over one year

The rate of ventricular/brain ratio change was significantly greater in patients than controls ($p<0.0001$) (Table 10-5). The sensitivity and specificity of ventricular-to-brain ratio to classify patients was 87% and 68% respectively.

Table 10-5. Annual percent change in ventricular-to-brain ratio at one year.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Mean % per year (SD)</th>
<th>Range</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>38</td>
<td>12.6 (6)</td>
<td>2.4 - 27</td>
<td>10.7 - 14.6</td>
</tr>
<tr>
<td>Controls</td>
<td>19</td>
<td>4.7 (4.1)</td>
<td>-2.3 - 14</td>
<td>2.7 - 6.7</td>
</tr>
</tbody>
</table>

Change over six months.

Over both six month periods, there was significant excess change in the ventricle-to-brain ratio in the patients compared to the controls ($p<0.0001$), and both the mean and standard deviation of the annualized rate of ventricular increase were similar to those seen over one-year (Table 10-6). However, the sensitivity and specificity were not as good: 76% and 75% for the first interval, and 71% and 63% for the second.
Table 10-6. Ventricle-to-brain ratio annual percent change over six month.

<table>
<thead>
<tr>
<th>Interval</th>
<th>Mean (SD) interval (days)</th>
<th>n</th>
<th>Mean % per year (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-6 months</td>
<td>180 (7)</td>
<td>38</td>
<td>11.7 (6.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>3.7 (6.6)</td>
</tr>
<tr>
<td>6-12 months</td>
<td>186 (14)</td>
<td>38</td>
<td>12.8 (8.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>19</td>
<td>5.7 (4.4)</td>
</tr>
</tbody>
</table>

**Power calculations**

Based on the assumptions above, using the ventricle-to-brain ratio change, at one year 131 patients would be required in each treatment arm. Based on the results of the first six month time period, this number would be 147 per arm; on those from the second: 215. Accounting for normal ageing, these results were 355 over one year, and 315 and 700 per arm over 6 months.

**10.4 Discussion**

The results of this work confirm that excess ventricular volume changes in patients compared to controls may be detected at intervals less than one year. This suggests that should a disease-modifying agent reduce the rate of ventricular expansion as a consequence of a reduction in brain atrophy, ventricular volume measures could be used as an outcome measure in six-month studies.

**Cross-sectional results**

The finding that baseline ventricular volumes are significantly increased in AD compared to controls has been demonstrated in several previous studies (Murphy et al, 1993; Wang et al, 2002; Jack et al, 1998). The results of the latter study, reporting baseline ventricular ranges of 11–82 ml in controls and 21–146 ml in AD, are similar to those in this study (15–113 ml in controls; 21–129 ml in patients). However, as has previously been reported, overlap between patients' and controls' volumes limits the use of ventricular volume alone as a diagnostic measure.
Longitudinal results

As has been previously demonstrated, change in ventricular volume better discriminated patients from controls than cross-sectional measures alone (Figure 10-2). This discriminative power over a one year period (sensitivity 85%, specificity 95%); was considerably reduced over a six-month interval. This unsurprisingly suggests that ventricular change is a better diagnostic marker when measured over longer periods. The excess ventricular expansion over time in AD demonstrated in this study has been reported in numerous previous studies. Some of these studies reporting ventricular increases in terms of ml/year are summarized in Table 10-7. These reports and particularly the latter studies performed using high resolution MRI demonstrate a broad consensus as to the magnitude of ventricular enlargement in AD. Discrepancies between these results are likely to be due to protocol differences and inter-subject variation.

The changes in AD are substantially higher than those seen in normal ageing. In this study, normal controls underwent ventricular change of 0.7-1ml/year, which compares with reports from Wang et al (Wang et al, 2002) who found 0.4ml per year, Schott et al in a study of controls in their 40s who found ventricular expansion of 0.44 ml per year (Schott et al, 2003), and Scahill et al who reported change of 0.65ml per year (Scahill et al, 2003). In this study there was significant evidence for ventricular expansion in control subjects, although there was not evidence for accelerating rates of ventricular expansion over time, possibly due to the small sample size and limited follow-up time. However Scahill et al (Scahill et al, 2003) found that rates of ventricular expansion increase with advancing age; this in part might account for the findings of Silbert et al who, despite reporting mean ventricular rates in AD that closely approximated to other studies, showed control subjects of average age 89 years undergoing ventricular change at 3.3ml per year. Another explanation might be comorbidity, as these subjects died within 2-3 year of their final measurements being made (Silbert et al, 2003).
Table 10-7. Reported absolute rates of ventricular expansion in AD (ml/yr).

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>CT/MRI</th>
<th>Rate/yr</th>
<th>n</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luxenberg</td>
<td>1987</td>
<td>CT</td>
<td>13 cc</td>
<td>18</td>
<td>1.3 yrs</td>
</tr>
<tr>
<td>(Luxenberg et al, 1987)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DeCarli</td>
<td>1992</td>
<td>CT</td>
<td>7.1 ml</td>
<td>20</td>
<td>3.9 yrs</td>
</tr>
<tr>
<td>(DeCarli et al, 1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shear</td>
<td>1995</td>
<td>CT</td>
<td>5.3 cc</td>
<td>24</td>
<td>2.1 yrs</td>
</tr>
<tr>
<td>(Shear et al, 1995)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang</td>
<td>2002</td>
<td>MRI</td>
<td>8.2 ml</td>
<td>14</td>
<td>1 yrs</td>
</tr>
<tr>
<td>(Wang et al, 2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silbert</td>
<td>2003</td>
<td>MRI</td>
<td>5.5 ml</td>
<td>24</td>
<td>5.8 yrs</td>
</tr>
<tr>
<td>(Silbert et al, 2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td>MRI</td>
<td>4.2 -5.7 ml</td>
<td>38</td>
<td>6 months</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 ml</td>
<td></td>
<td>1 year</td>
</tr>
</tbody>
</table>

Another means of calculating ventricular enlargement is as a percentage, either of brain volume, or of total intracranial volume. Theoretically, by providing a protection against inter-individual differences such measurements might be expected to provide more consistent results. Results from studies adopting this approach are summarized in Table 10-8. Despite different methods of calculating this ratio, ventricle-to-brain ratios do show remarkably consistent changes over time, all reports using high resolution MRI showing percentage ventricular change of approximately 12-14 percent per annum.
Table 10-8. Reported annual percentage increases in ventricular size in AD.

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>CT/MRI</th>
<th>Rate/yr</th>
<th>n</th>
<th>Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Leon (de Leon et al, 1989)</td>
<td>1989</td>
<td>CT</td>
<td>9%</td>
<td>50</td>
<td>3yrs</td>
</tr>
<tr>
<td>Jack* (Jack et al, 1998)</td>
<td>1998</td>
<td>MRI</td>
<td>14%</td>
<td>24</td>
<td>1.89yr</td>
</tr>
<tr>
<td>Wang (Wang et al, 2002)</td>
<td>2002</td>
<td>MRI</td>
<td>13.2%</td>
<td>14</td>
<td>1yrs</td>
</tr>
<tr>
<td>Bradley (Bradley et al, 2002)</td>
<td>2002</td>
<td>MRI</td>
<td>13%</td>
<td>5</td>
<td>Multiple intervals, 2.5-7 months</td>
</tr>
<tr>
<td>This study</td>
<td></td>
<td></td>
<td>11.7-12.8%</td>
<td>38</td>
<td>6 months 1year</td>
</tr>
</tbody>
</table>

*: Temporal horn of lateral ventricles only

It is notable that in this study, ventricle-to-brain ratio change performed worse as a diagnostic test than did ventricular change alone, demonstrating that correcting for brain volume does not always improve group separation. This is likely to reflect the inherent stability of measures of ventricular volume: the ventricular system is centred deep within the brain and is therefore relatively protected from changes in position and thus from inhomogeneity. Furthermore, the ventricular-brain-boundary is a high contrast boundary, increasing ease of segmentation. The brain as a whole is a larger structure and is more susceptible to inhomogeneity artefact, and edge/field of view effects.

The sample size estimates in this study suggest that trials using ventricular size as an outcome measure may be feasible at intervals of less than one year. Sample size estimates were very similar using either ventricular change or ventricle-to-brain ratio change over one year, although, when correction for the effect of normal ageing was made, the ratio performed less well, due to the increased average rate of change in this index compared to ventricular volume change alone.
There was some variability in the results from the two six-month intervals. Thus for the first interval, the sample size estimates were better using ventricular volumes alone, and for the second, the ratio performed better. However, when correcting for the effects of normal ageing, the ventricle-to-brain ratio again performed considerably worse. Sample size estimates were higher in the second six months than for the first. This is a function of the increased variability seen in the patients over the second period, as power estimates are largely driven by the variance of the patient group. This increased variance is likely to reflect both random variability, as well as the more variable length of follow-up in the second six month period.

The sample size estimates of 147-200 per treatment arm at six months are similar to those reported by Bradley et al who calculated, based on five patients undergoing several scans over a 2.5 to 7 month interval, that 135 subjects would be needed per treatment arm for a six-month trial, using a ventricle-to-brain ratio as an outcome measure (Bradley et al, 2002). Such estimates compare favourably with estimates of whole brain atrophy at one year e.g. 115 per arm, using the brain boundary shift integral (Fox et al, 2000). Furthermore, as drop-out is likely to be less over six months than one year, sample size estimates may benefit further from short interval design. It is of note that the sample sizes are not dramatically increased when the interval is reduced from one year to six months. This is a consequence of very similar standard deviations of the results, suggesting a high degree of consistency in ventricular changes both within and between individuals.

Ventricular measures over short intervals are therefore potential candidates as progression markers for AD. Nonetheless, other factors need to be taken into account when choosing a trial outcome measure. Ventricular measurements require manual outlining by trained raters. Although such measurements are considerably quicker to learn and perform and are less variable than hippocampal or entorhinal cortex measurements (Schott et al, 2003), there are both time and financial implications for their use. Furthermore questions still remain as to the suitability of such measures as sole surrogate markers for AD. Silbert et al demonstrated that ventricular change correlated with accumulation of neurofibrillary tangles and senile plaques at post-mortem, but also commented that AD neuropathology did not fully explain the volumetric changes and suggested that volumetric brain changes were a parallel and
indirect correlated marker of AD pathology (Silbert et al, 2003). It follows that substantial change in ventricular size could occur without affecting the histopathological features of AD. Other factors known to alter cerebrospinal fluid volumes include treatment with diuretics, dialysis (Walters et al, 2001), stage of the menstrual cycle (Teasdale et al, 1988) and hypercapnia (Teasdale et al, 1988). Zipursky et al demonstrated reversibility in ventricular enlargement in less than a month in 10 chronic alcoholics who abstained from alcohol (Zipursky et al, 1989). In all these cases, change in ventricular size (that can be seen over short periods) is clearly unrelated to AD.

Assuming ventricular dilation is a biological marker for AD, this study suggests that ventricular change may be a useful diagnostic marker to distinguish AD from controls over a one year period. However, it is unlikely to be helpful in the differential diagnosis of dementia syndromes. Either ventricular changes or change in ventricle-to-brain ratio may be useful as outcome measures for AD treatment trials over short intervals (e.g. six months), although the latter may be less useful if the effect of normal ageing is to be controlled for. However, to minimize the possible influences of other factors on ventricular volume, care is needed to monitor the use of other therapies and comorbidity during any such trial period, and additional direct measures of brain atrophy are likely to be required to ensure that any changes are truly due to reduction in cerebral atrophy.
11 The brain boundary shift integral

11.1 Introduction

The results presented in the previous chapter demonstrate that ventricular change may be an effective way of tracking the progression of AD, but may have limitations as a surrogate marker. Direct measures of brain volume or atrophy, by being “closer” to the intended effect of a drug (i.e. halting or slowing the build-up of the histopathological features of AD leading to a reduction in neural death) may be a better surrogate marker of AD progression. Furthermore, such direct measures may be less likely to be influenced by many of the factors previously discussed. Whilst physiological variability is inherent in any biological system, and therapeutic interventions could possibly alter brain volume through effects of hydration or swelling, it is perhaps more likely that regulatory agencies such as the FDA would approve a drug as having disease-modifying effects if it were to alter direct rather than indirect measures of brain atrophy.

Several methods of measuring whole brain atrophy from serially acquired volumetric brain scans have been developed. The majority rely on positional matching of serially acquired scans (registration), followed by a quantification process. At its simplest level, quantification can be achieved by simple subtraction of the volumes of the registered brains from one another. For this to be feasible, either scans must be corrected for total intracranial volume (Edland et al, 2002; Whitwell et al, 2001) or registered using a 9dof registration (see chapter 5) to provide protection against voxel drifts. Such subtraction measures require individual measurements at each time point, inevitably increasing measurement error.

Measuring cerebral atrophy from serial MRI scans requires accurate and robust techniques capable of analysing large numbers of scans acquired from multiple centres; ideally these analyses would be done by reliable semi- and/or fully automated image analysis techniques. One assessment of the performance of these techniques involves calculating how well they distinguish atrophy rates in AD from normal ageing (Gunter et al, 2003). The boundary shift integral (BBSI) has been proposed as a measure of cerebral volume change from serial MRI (Gunter et al, 2003;
Freeborough and Fox, 1997). This technique is based on identifying the brain tissue in two different scans of the same individual, registering the two scans spatially, and then subtracting scan intensities in the area between the borders of the aligned brains. Once the two scans have been segmented, the remainder of the process is fully automated, requiring only two user defined parameters that define the location and width of the sampling (image intensity) window (see Figure 11-1). The BBSI increases the precision of estimation of atrophy (as compared to the differences in segmented volumes) by eliminating some random measurement error. However, it also introduces a systematic underestimate of volume change due to some of the changes in scan intensity not spanning the parameter-defined window. Previous studies have chosen values for an entire study for these parameters based on analysis of a single scan (Freeborough and Fox, 1997) or selected them from the outcomes of scan normalization processes (Gunter et al, 2003).

In this chapter the effect of varying the two windowing parameters on atrophy measurements was assessed using scans from a subset of patients with AD and age matched controls. Agreement between BBSIs and segmented volume differences (SVDs) was critically assessed, and an improved set of BBSI parameters derived. The clinical utility of this process was assessed by using these new parameters to calculate the atrophy rates in patients and age matched controls and to assess the separation in atrophy rates of these groups, and the approximation of the mean derived atrophy rates to those calculated from segmented volume differences.

11.2 Methods

Scans from 23 patients with AD and 12 controls were included in this study. Scans were acquired using the previously defined parameters approximately one year apart (visits 1 and 7). As previously described: (a) both baseline and repeat scans were segmented to classify brain tissue; (b) the repeat scan was registered to the baseline using a 9dof registration; (c) chirp-z interpolation was used to resample the intensities of the registered repeat scan to the baseline.
11.2.1 The boundary shift integral: theoretical aspects

The BBSI operates on the following principals. Viewed in cross-section, the boundary between brain and cerebrospinal fluid (CSF) may be seen as an intensity curve, crossing between brain (bright) and CSF (dark) signals. This baseline profile $i_{base}(x)$ can be considered as a continuous line. The corresponding intensity profile for the second scan $i_{reg}(x)$, accurately registered to the first registered scan, is seen to similarly start in brain tissue and curve to CSF. If the boundary has shifted by an amount $\Delta w$ in the interval between the scans, then the profiles may be depicted as in Figure 11-1.

Figure 11–1. The brain boundary shift integral
A representation of an intensity profile through a boundary of the brain on both a baseline scan and a registered repeat scan; labelled as $i_{base}(x)$ and $i_{reg}(x)$ respectively. During the interval between the scans the boundary had shifted by an amount $\Delta w$, which we determine as the area A divided by (11-2).

![Figure 11-1](image_url)

The area labelled A in the figure represents the integral with respect to intensity of the difference in position of the profiles over the interval $[I_1, I_2]$; it follows that

$$A = (I_1-I_2) \Delta w \quad (1)$$
The area $A$ may alternatively be calculated as integral of the intensity profile of the registered scan from that of the baseline over the range $I_1 - I_2$

$$A = \int_{\text{brain boundary}} \left( \text{clip}(i_{\text{base}}(x), I_1, I_2) - \text{clip}(i_{\text{reg}}(x), I_1, I_2) \right) dx$$  \hspace{1cm} (2)

where

$$\text{clip}(a, I_1, I_2) = \begin{cases} 
I_1 & a < I_1 \\
I_2 & a > I_2 \\
a & I_1 \leq a \leq I_2 
\end{cases}$$

Thus $\Delta w$, the boundary shift may be calculated by equating (1) and (2),

$$\Delta w = \frac{1}{I_1 - I_2} \int_{\text{brain boundary}} \left( \text{clip}(i_{\text{base}}(x), I_1, I_2) - \text{clip}(i_{\text{reg}}(x), I_1, I_2) \right) dx$$  \hspace{1cm} (3)

Extending this analysis to three dimensions, and approximating the analytic integral by a numerical integral computed over the sampling intervals of the MR data, we obtain, the \textit{brain boundary shift integral} (BBSI):

$$\text{BBSI} = \frac{K}{I_1 - I_2} \sum_{x,y,z \text{brain region}} \left( \text{clip}(i_{\text{base}}(x,y,z), I_1, I_2) - \text{clip}(i_{\text{reg}}(x,y,z), I_1, I_2) \right)$$  \hspace{1cm} (4)

where $K$ is the product of the sampling intervals in each dimension (i.e. the voxel volume). The BBSI therefore represents the total volume traversed by the boundaries of the brain going from the first scan to the second scan, and is a direct measure of change in volume.

Evaluation of the BBSI requires the selection of appropriate values for $I_1$ and $I_2$; both values should be within the range of the intensity transitions at the boundaries of the brain. Using $I_1$ and $I_2$, a window size (the difference between $I_2$ and $I_1$), $Iw$, and window centre (the average of $I_2$ and $I_1$), $Ic$ can be calculated; $Ic$ and $Iw$ are the required user defined input parameters to the algorithm.
11.2.2 Experiments

The BBSI was calculated for all 35 subjects in this section of the study over a varying set of \( I_c \) and \( I_w \). As CSF typically has a normalised intensity below 0.35 (i.e., < 35% of the mean intensity of the interior region), whilst grey matter intensity is typically ~0.8, a variety of window centres over the range from 0.4 to 0.75 in intervals of 0.05 were assessed; \( I_w \) was varied from 0.05 to 0.5 in intervals of 0.05. Using the chosen parameters, changes in whole-brain volume were calculated using the BSI, and expressed as an annualised percentage change in brain volume from baseline, assuming a constant proportionate rate of brain loss. These were then compared both with the segmented volume differences and between patients and controls.

11.2.3 Statistical methods

In order to determine improved values of \( I_c \) and \( I_w \), \( t \)-tests were used to assess the difference between volume changes estimated from the segmented volume differences and BBSIs. Pitman’s test was used to assess the difference in variance of these estimates for a range of values of \( I_c \) and \( I_w \). If the BBSI reduces variability without introducing material bias, little difference would be expected between the volume change calculated from the SVD and that estimated using the BBSI, but the variability of BBSI derived changes should be smaller. A \( t \) or \( z \) statistic greater than 3 (equating to an unadjusted p-value of the order of 0.001) was regarded as providing evidence of a difference in means or variances as appropriate. Thus the smaller the \( t \) statistic, the better the approximation of the BBSI to the SVD; and the larger the \( z \) statistic the smaller the variance in the BBSI. An unpaired \( t \)-test with unequal variances was used to compare the change in atrophy rates for a change in \( I_c \) and \( I_w \) in patients with the equivalent change in controls as compared to the original \( I_c \) and \( I_w \) of (0.5, 0.5).

11.2.4 Results

Calculations of the \( t \)-statistic for a range of \( I_c \) and \( I_w \) are shown in Figure 11-2, with \( z \)-statistics shown in Figure 11-3. Figure 11-2 demonstrates that the difference between the distributions of the means of the SVDs and the BBSIs was minimised at an \( I_c \) of ~0.65. The \( t \)-statistic reduces in value as \( I_w \) is increased for values of \( I_c \) of 0.6,
but the evidence for a difference is not great up to an $l_w$ of 0.2. Figure 11-3 demonstrates that the difference in variance is greatest at the extremes of $l_c$ when $l_w$ is large, and in the raised area of the curves ($l_c = 0.55$ to 0.6) for smaller values of $l_w$.

As a compromise between the increased robustness achieved by ensuring a reasonably wide $l_w$ (meaning more voxels contribute to the BBSI equation) and a small loss in accuracy, $l_c=0.6$, $l_w=0.2$ were selected as the new parameters. In comparison with the “standard parameters” ($l_c=0.5$, $l_w=0.5$) the $t$-statistic comparing means was substantially reduced from over 2 to less than 1, representing a increase in the accuracy of the BBSI. The $z$-statistic was reduced marginally from approximately 3.0 to 2.5, demonstrating a small reduction in the variance of the BBSI compared to the SVD.

Figure 11–2. $t$-statistics for the mean difference between BBSI and SVD

The lower the $t$ statistic, the smaller the difference between the BBSI and the SVD, i.e. the better the approximation of BBSI to “true atrophy”
Figure 11–3. z-statistics from a range of BBSI and SVD values

The larger the z-statistic, the greater the difference in variance between the BBSI and SVD; as the variance of the SVD is unchanged, this represents a reduction in the variance of the BBSI.

Changing the parameters from $I_c=0.5$, $I_w=0.5$ to $I_c=0.6$, $I_w=0.2$ increased the estimated mean atrophy rates in both groups so that they were closer to the segmented volume differences (Table 11–1) with some evidence ($p = 0.06$) that the change was greater in patients than controls and therefore that separation between the patient and control groups was increased. The standard deviations of the BBSI settings were lower than those of the segmented volume differences, demonstrating the decreased variability associated with the direct (BBSI) calculation.
Table 11-1. BBSI derived atrophy vs. segmented volume differences.

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD) annual percent atrophy rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New BBSI settings $I_c=0.6$, $I_w=0.2$</td>
</tr>
<tr>
<td>AD</td>
<td>2.27 (0.83)</td>
</tr>
<tr>
<td>Control</td>
<td>0.5 (0.46)</td>
</tr>
</tbody>
</table>

11.3 Discussion

These results suggest that by altering the parameters of the BBSI, it is possible to increase the accuracy of whole brain atrophy measurements without reducing the precision of the method. The alteration in atrophy rates derived using this method is likely to have particularly significance in clinical trials. As power is determined by variance, decreasing the standard deviation of the measurement will decrease the estimated sample sizes required in a clinical trial. Furthermore, altering these parameters, the BBSI derived atrophy rates are seen to more closely approximate to those from the segmented volume differences, whilst reducing the inherent (and, in a clinical trial, punitive) errors associated with the latter.

This optimization technique can be easily implemented to any serial volumetric MRI data set, and by basing parameter selection on the entire data set rather than an individual scan, the chosen parameters are less likely to be influenced by incorrect segmentations than approaches using one or only a few scans, as originally described (Freeborough and Fox, 1997). Finally, this method could be utilised in serial MRI trials investigating a patient group in the absence of a control arm, unlike methods that use separation of patients and controls as the gold standard for determining parameters.
12 Atrophy over short intervals using the boundary shift integral

12.1 Introduction

Few studies have assessed the potential of global brain atrophy as a potential outcome measures in AD over intervals of less than one year. Having determined improved BBSI settings, this technique was used to calculate whole brain atrophy rates over a variety of short intervals. Firstly, the BBSI atrophy rates were calculated for the same cohort for whom ventricular rates were calculated, to allow for direct comparison. Secondly, using the best-rated of the two initial (visit one) scans as a baseline, atrophy rates were determined prospectively over intervals of 2 weeks, 6 weeks, 3 months, 6 months, 9 month, 12 months, 18 months and two years. An assessment of the robustness of the BBSI to scan quality was performed by comparing the estimated sample size required at each interval when all, and when just “good” scans (see chapter 3) were included in the analysis.

12.1.1 Methods

Changes in whole-brain volume were calculated directly from the registered images using the BBSI (with the “optimised” parameters outlined in chapter 11), and expressed as an annualised percentage change in brain volume from baseline, assuming a constant proportionate rate of brain. To allow comparison with future work, BBSI calculations were carried out on log scales. Atrophy rates were compared between patient and controls using t-tests.

In order to compare ventricular and BBSI derived atrophy rates and power calculations over six months and one year, the same cohorts as described in chapter 6 were assessed. BBSI derived atrophy rates were calculated over two time periods: one year, and six months (visit 1 – visit 5). To assess change over a variety of intervals in a prospective manner, atrophy rates were calculated starting from the best rated of the two baseline (visit 1) scans, to scans acquired at each of the other visits (visits 2-7; 9-10). Where two scans were performed on the same day (visits 3, 6) the best rated (see chapter six) of these two scans was used.
Power calculations were carried out to estimate numbers needed in each arm to power a placebo-controlled study to detect a 20% reduction in rate of atrophy (90% power, two-tailed 5% significance, not accounting for normal ageing) at each time point. For the multiple time point analysis, power estimates were also derived to include the effect of normal ageing (taken as the mean calculated over the longest period of follow-up i.e. 0.46% per annum) and scan drop-out (i.e. multiplied by number of scans included in the complete analysis/number of scans included in “good” only analysis).

12.2 Results

12.2.1 Ventricular expansion vs. BBSI derived brain atrophy rates

The cohort demographics are described in chapter 6. The BBSI derived annual percentage atrophy rates are shown in Table 12-1, alongside the previously described annualized ventricular change. Power calculations suggest that to demonstrate a 20% reduction in mean atrophy rate in the patients, for each arm of a placebo-control study (90% power, two-tailed 5% significance, not accounting for normal ageing), at six months: 170 patients would be required using ventricular measures, and 401 using the BBSI. At one year, 134 would be required using ventricular measures, and 159 per arm using the BBSI. Atrophy rates were able to distinguish patients from controls at one year with a sensitivity of 84% and a specificity of 89% using the BBSI, and a sensitivity of 85% and specificity of 95% using ventricular volumes. At six months, the sensitivity and specificity were: 81% and 80% using the BBSI; and 87% and 85% using ventricular volume change.
Table 12-1 Ventricular change and BBSI-derived change over 6 and 12 months
Comparison of annualized ventricular change (ml) and BBSI derived change (percent) in the same cohort over periods of six months and one year.

<table>
<thead>
<tr>
<th>Interval</th>
<th>n</th>
<th>Mean (SD) change/yr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ventricle change</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One year</td>
<td>Patients 38</td>
<td>4.95 (2.5) ml</td>
</tr>
<tr>
<td></td>
<td>Controls 19</td>
<td>1.04 (1.1) ml</td>
</tr>
<tr>
<td>Six months</td>
<td>Patients 38</td>
<td>4.22 (2.4) ml</td>
</tr>
<tr>
<td></td>
<td>Controls 19</td>
<td>0.78 (1.3) ml</td>
</tr>
<tr>
<td><strong>BBSI</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One year</td>
<td>Patients 38</td>
<td>2.16 (1.1) %</td>
</tr>
<tr>
<td></td>
<td>Controls 19</td>
<td>0.73 (0.4) %</td>
</tr>
<tr>
<td>Six months</td>
<td>Patients 38</td>
<td>2.07 (1.8) %</td>
</tr>
<tr>
<td></td>
<td>Controls 19</td>
<td>0.56 (1.0) %</td>
</tr>
</tbody>
</table>

12.2.2 BBSI derived atrophy rates over multiple scan intervals from baseline

BBSI derived atrophy rates from “best” baseline for both patients and controls using all available scans are shown in Table 12-2. The standard deviation of mean atrophy rates can be seen to fall dramatically with increasing intervals from 2 weeks to six months. Thereafter the SD continues to decline slowly as the interval increases. The mean rate of atrophy over the longest interval was approximately 2.2% per annum for patients, and ~0.46% per annum for controls.

The results for the same intervals, this time excluding any scan rated as “bad” (see chapter 3) are shown in Table 12-3. Results from the two analyses were very similar, with no material differences in mean or variance for any interval. This is reflected in the sample size estimates shown in Table 12-4. No consistent difference in estimated sample size was seen whether or not only “good” scans were included in the analysis. When sample sizes were corrected for the percentage of scans excluded in the “good” only analysis, the estimates were consistently higher in the analysis only using “good” scans.
Table 12-2. BBSI derived atrophy over multiple intervals from baseline (all scans)

"Best scan" only used at time points with two scans (V1, V3, V6)

<table>
<thead>
<tr>
<th>Visit</th>
<th>Interval (days)</th>
<th>Alzheimer's disease atrophy rate (%/yr)</th>
<th>Control atrophy rate (%/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start - end</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>V1-V2</td>
<td>16.0 (5)</td>
<td>2.86 (18)</td>
<td>43</td>
</tr>
<tr>
<td>V1-V3</td>
<td>62 (5)</td>
<td>2.45 (6.3)</td>
<td>41</td>
</tr>
<tr>
<td>V1-V4</td>
<td>98 (8)</td>
<td>1.82 (3.4)</td>
<td>44</td>
</tr>
<tr>
<td>V1-V5</td>
<td>180 (7)</td>
<td>2.12 (1.7)*</td>
<td>42</td>
</tr>
<tr>
<td>V1-V6</td>
<td>270 (18)</td>
<td>2.26 (1.8)*</td>
<td>37</td>
</tr>
<tr>
<td>V1-V7</td>
<td>365 (14)</td>
<td>2.32 (1.5)*</td>
<td>42</td>
</tr>
<tr>
<td>V1-V9</td>
<td>551 (17)</td>
<td>2.35 (1.3)*</td>
<td>26</td>
</tr>
<tr>
<td>V1-V10</td>
<td>730 (11)</td>
<td>2.19 (0.9)*</td>
<td>13</td>
</tr>
</tbody>
</table>

*Significant difference between annual rate of atrophy comparing patients and controls (p<0.0001, t-test)
Table 12-3. Atrophy rates calculated using the BBSI and "good" scans only. "Best scan" only used at time points with two scans (V1, V3, V6)

<table>
<thead>
<tr>
<th>Visits</th>
<th>Interval days</th>
<th>Alzheimer's disease</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) days</td>
<td>Mean (SD) Atrophy rate</td>
<td>n</td>
</tr>
<tr>
<td>V1-V2</td>
<td>16 (4)</td>
<td>1.26 (17.3)</td>
<td>35</td>
</tr>
<tr>
<td>V1-V3</td>
<td>53 (5)</td>
<td>1.54 (5.6)</td>
<td>35</td>
</tr>
<tr>
<td>V1-V4</td>
<td>98 (9)</td>
<td>1.55 (2.55)</td>
<td>37</td>
</tr>
<tr>
<td>V1-V5</td>
<td>180 (8)</td>
<td>2.12 (1.7)*</td>
<td>36</td>
</tr>
<tr>
<td>V1-V6</td>
<td>270 (16)</td>
<td>2.22 (1.8)*</td>
<td>36</td>
</tr>
<tr>
<td>V1-V7</td>
<td>365 (15)</td>
<td>2.30 (1.5)*</td>
<td>40</td>
</tr>
<tr>
<td>V1-V9</td>
<td>551 (16)</td>
<td>2.38 (1.4)*</td>
<td>25</td>
</tr>
<tr>
<td>V1-V10</td>
<td>728 (9)</td>
<td>2.18 (0.9)*</td>
<td>13</td>
</tr>
</tbody>
</table>

*Significant difference between annual rate of atrophy comparing patients and controls (p<0.0001, t-test)
Table 12-4. Sample size estimates for a variety of scan intervals from baseline. Estimates are made using the BBSI derived atrophy rate from: all scans with and without including the effect of normal ageing; and “good” scans only, with and without including the effect of normal ageing, including adjustment for the percentage of scans “lost” due to the exclusion of poor scans.

<table>
<thead>
<tr>
<th></th>
<th>All scans</th>
<th>All scans + Normal ageing*</th>
<th>Good scans</th>
<th>Good scans + Normal ageing*</th>
<th>Good scans + Normal ageing* + lost scans</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 weeks</td>
<td>20791</td>
<td>29525</td>
<td>98949</td>
<td>245455</td>
<td>301559</td>
</tr>
<tr>
<td>6 weeks</td>
<td>3471</td>
<td>5261</td>
<td>6941</td>
<td>14112</td>
<td>16531</td>
</tr>
<tr>
<td>3 months</td>
<td>1832</td>
<td>3281</td>
<td>1421</td>
<td>2873</td>
<td>3417</td>
</tr>
<tr>
<td>6 months</td>
<td>338</td>
<td>550</td>
<td>338</td>
<td>550</td>
<td>642</td>
</tr>
<tr>
<td>9 months</td>
<td>333</td>
<td>525</td>
<td>345</td>
<td>549</td>
<td>564</td>
</tr>
<tr>
<td>12 months</td>
<td>219</td>
<td>314</td>
<td>223</td>
<td>349</td>
<td>366</td>
</tr>
<tr>
<td>18 months</td>
<td>161</td>
<td>248</td>
<td>182</td>
<td>279</td>
<td>290</td>
</tr>
<tr>
<td>24 months</td>
<td>89</td>
<td>142</td>
<td>89</td>
<td>144</td>
<td>144</td>
</tr>
</tbody>
</table>

*: rate of normal ageing taken as 0.46% /year (i.e. result from longest interval).
12.3 Discussion

12.3.1 Ventricular and BBSI derived atrophy rates

At one year, when the BBSI derived atrophy rates and ventricle size changes were compared on the basis of sample size estimates, there was little difference between the two measures. This can be explained as the ratio of mean-to-standard-deviation (the basis of power estimates) was very similar between the groups (1.98, ventricles: 1.82 BBSI). However at six months, the estimated sample sizes based on the BBSI were more than double those of the ventricular measure; this is reflected by the much lower mean-to-standard-deviation ratio (1.14 vs. 1.76). This would suggest that ventricular measures might be more sensitive or reliable surrogates of AD progression over short intervals than direct measures of brain volume, which are considerably more variable. Similarly the improvement in specificity at six months would suggest that at short intervals ventricular measures may be better diagnostic markers than BBSI derived atrophy rates. As discussed previously, these potential advantages must be offset against the possibly less robust nature of ventricles as true surrogates of AD progression, and the increased time required undertaking such manual measures. This may be particularly so if these measures are used for diagnostic purposes.

12.3.2 BBSI derived atrophy rates over multiple scan intervals from baseline

This study is the first to assess atrophy rates derived over multiple short scan intervals in a prospective cohort of AD using the BBSI. The results suggest that it is possible to achieve a realistic estimate of the mean rate of atrophy in such a cohort over periods as short as three months, although six months were required before significant differences between groups of patients and controls could be determined given the group sizes assessed. However, the variance (or standard deviation) of the group, is the most important factor when using such measures in clinical trials. The standard deviation of the atrophy measures in the controls is seen to decline sharply over the first six months; from then on, there is a slower but still important decline right up to two years. The very large standard deviation seen at short intervals is likely to be a combination of inherent variability in atrophy over short intervals, and the problems of extrapolating changes (and hence errors) to a one year atrophy rate (i.e. a small
denominator such as two weeks means errors have a very large effect on annualized rates); it is likely that the latter effect is the more important. These results suggest that a drug trial using global brain atrophy measured over two time points as an outcome measure should use two time points separated over as long a period as possible. However, in practice, this strategy must be weighed up against the costs and time pressures in running studies, the likely increased drop-out rates with longer designs, and the probable intra-subject acceleration of atrophy (Chan et al, 2003) that occur over longer periods.

In controls, the atrophy rate varied between 0.46–0.62% per annum at intervals over six months. Again, the SD continued to decline up to periods of 18 months. This estimate of brain atrophy in normal ageing is in line with previous studies using a variety of techniques (for review see (Fox and Schott, 2004)). This study further confirms the BBSI as a robust measure of atrophy (Freeborough and Fox, 1997). Whether all or just "good" scans were used in the analysis, the results were essentially unchanged; the cost of missed data points in a clinical trial is high, suggesting that using all available scans for analysis is the optimal strategy.

Whilst few studies have assessed atrophy rates at short intervals, several studies have reported annualized atrophy rates, generally assessed from studies performed over a number of years. Some of these studies assessing whole brain atrophy are summarized in Table 12-5.
Table 12-5. Reported MRI brain atrophy rates and power calculations in AD
Power calculations assuming 20% reduction, no drop-outs or unusable scans, not accounting for normal ageing, 2 tailed, 5% significance

<table>
<thead>
<tr>
<th>First author</th>
<th>Year</th>
<th>Structure</th>
<th>Mean (SD) Rate %/yr</th>
<th>n</th>
<th>Estimated sample size per arm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fox (Fox et al, 1999a)</td>
<td>1999</td>
<td>Brain*</td>
<td>2.4 (1.4)</td>
<td>29</td>
<td>179</td>
</tr>
<tr>
<td>Fox (Fox et al, 2000)</td>
<td>2000</td>
<td>Brain*</td>
<td>2.37 (1.1)</td>
<td>15</td>
<td>113</td>
</tr>
<tr>
<td>O'Brien (O'Brien et al, 2001)</td>
<td>2001</td>
<td>Brain*</td>
<td>2.0 (0.9)</td>
<td>9</td>
<td>106</td>
</tr>
<tr>
<td>Bradley (Bradley et al, 2002)</td>
<td>2002</td>
<td>Brain</td>
<td>2.14 (0.5)</td>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>Wang (Wang et al, 2002)</td>
<td>2002</td>
<td>Brain</td>
<td>2.4 (1.2)</td>
<td>14</td>
<td>131</td>
</tr>
</tbody>
</table>

*: Using the BBSI

In this study, the one year atrophy rate in AD was calculated to be 2.32%/yr, with an SD of 1.5%/yr. Of the three other studies performed using the BBSI, the mean can be seen to be similar, but the SD in this study appears to be slightly higher. The sample size in this study is considerably larger than those previously reported, and the study was carried out prospectively in an elderly population of patients with varying stages of AD, all factors that may account for this increased variability. Some degree of variability is also likely to be due to random error. It is notable that a smaller sample of patients from the same cohort (those used in the comparison with the ventricular results), in whom different scan pairs were assessed, produced a slightly lower mean and standard deviation. The differences in estimated sample size (159 – 219, an increase of nearly 40%) demonstrate the profound effects such random errors can have on power calculations which are critically dependent on variances.
Given the variability of results within the same cohort using the same imaging protocols, intervals and analysis techniques, it is clearly more difficult to compare these results with studies using different methodologies. Thus Bradley et al, who report a similar mean atrophy rate with a much reduced SD calculated this using a linear mixed model, rather than a direct forward comparison in 5 patients each having a number of scans over short periods (Bradley et al, 2002). Nonetheless, despite large differences in variance between studies, the mean rates of atrophy in all the studies are remarkably similar.

With the exception of Bradley et al (Bradley et al, 2002) few studies have addressed rates of whole brain atrophy over periods of less than one year, and none has used the BBSI. The results of this study suggest that statistically significant differences between patients and controls can be established at intervals as short as six months, and that at this interval sample size estimates begin to be practically feasible for use in trials.

Whilst the cohorts examined over intervals of up to one year were more or less identical, by virtue of the smaller numbers continuing for the additional section of the trial, the results over the 18 month and two year intervals may not be directly comparable. A more robust and accurate means of determining true sample size requirements would be to determine true rates of atrophy based not on single scan pairs, but on atrophy derived from multiple scan intervals. Given that variance of atrophy is the key determinant of power, if strategies to reduce such variability could be implemented, trials over shorter periods might be more feasible. Such strategies might include combining multiple scans and scan intervals, or sub-stratifying the AD population into more homogeneous groups. A robust mathematical framework for modelling atrophy would allow these hypotheses to be tested: these aims were the motivation for the work described in the following chapters.
13 Modelling atrophy

The results from the previous chapter demonstrate that even when the subgroups from the same patient cohort are assessed at different time points with the same interval between scans using the same methodology, variable estimates for mean (and SD) brain atrophy rate can be derived. The differences in mean, and particularly in variance, have profound implications for trials using such measures as progression markers. The aim of the work in this chapter was: (1) to introduce a multi-level model designed specifically to model these data; (2) to use this model to predict the underlying true mean and variance of atrophy in this cohort using the BBSI, and the constituents of this variance; and (3) to validate the model for these data. The parameters thus designed could then be taken forward to estimate the underlying mean and SD of atrophy with different trial designs and the associated sample size estimates.

13.1 Multiple intervals – multiple measurements

The data presented so far in this thesis have calculated atrophy from one given time point, i.e. the baseline scan through to another given time point. However, as a consequence of taking multiple scans at a number of time points, more than one measurement could be made for a given interval. Thus, for example, seven different scan pairs can be used to calculate atrophy over approximately a six month interval. These intervals, and the associated mean, SD and sample size estimates (not accounting for normal ageing, or dropouts) calculated using the BBSI are shown in Table 13-1.
Table 13-1. Atrophy over several different six month intervals.

Mean (SD) annual atrophy rates for patients with AD calculated using the BBSI, and associated samples size estimates (for a 20% effect, 90% power, 5% significance, two tailed) are shown for a number of approximately six month intervals [V: Visit; S: Scan number, if >1 performed at one visit]

<table>
<thead>
<tr>
<th>Interval</th>
<th>Mean (SD) interval (days)</th>
<th>n</th>
<th>Mean (SD) atrophy rate</th>
<th>Sample size estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI S1 to V5</td>
<td>180 (7)</td>
<td>43</td>
<td>2.09 (1.7)</td>
<td>347</td>
</tr>
<tr>
<td>VI S2 to V5</td>
<td>180 (7)</td>
<td>41</td>
<td>1.93 (1.4)</td>
<td>276</td>
</tr>
<tr>
<td>V4 to V6 S1</td>
<td>172 (20)</td>
<td>38</td>
<td>2.30 (2.2)</td>
<td>480</td>
</tr>
<tr>
<td>V4 to V6, S2</td>
<td>171 (19)</td>
<td>35</td>
<td>2.22 (2.4)</td>
<td>613</td>
</tr>
<tr>
<td>V5 to V7</td>
<td>185 (14)</td>
<td>43</td>
<td>2.47 (1.8)</td>
<td>279</td>
</tr>
<tr>
<td>V7 to V9</td>
<td>189 (20)</td>
<td>26</td>
<td>2.34 (1.5)</td>
<td>216</td>
</tr>
<tr>
<td>V9 to V10</td>
<td>172 (28)</td>
<td>13</td>
<td>2.38 (2.2)</td>
<td>448</td>
</tr>
</tbody>
</table>

This table demonstrates that widely differing results can be obtained prospectively from the same overall cohort scanned over virtually the same interval. Some variability is inevitable given the inherent imprecision in the measurements, and a certain amount of true physiological variability in brain volumes. The factors that influence this variability are also likely to include the differing number of observations (and hence different individuals being involved in each calculation), the variability (seen as SD) in the follow-up time, as well as random error. As statistical power \( \sim 1/\text{variance} \) (i.e. \( 1/\text{SD}^2 \)), it is clear that the variability in mean atrophy rates
has profound implications for sample size estimates: for trials over six months, the estimated number per treatment arm varies from 216 to 613, nearly a 200% difference.

This study design allows for the calculation of multiple atrophy rates over different time periods for each individual, as each acquired scan can be compared to each other acquired scan. Thus an individual completing visits 1 – 7 (i.e. one year), would have completed 10 different scans; comparing each scan to every other scan allows 45 BBSI measures of atrophy to be made. For an individual attending all 10 visits (two years), 12 scans would be acquired, and thus 66 measures of atrophy could be made.

Thus for the entire cohort atrophy rate calculations can be performed over a large number of different intervals ranging from intervals of 0 weeks (same day scans) to 2 years.

13.1.1 Methods

Every scan from each individual in the cohort previously described was registered to every other scan from the same individual, using a 9dof registration. The BBSI (as optimized in chapter 11) was used to calculate the volume of change on a log scale, which was then converted to an mean and SD annual rate of brain atrophy for each time point, both for patients and controls.

13.1.2 Results

For the patients, 2199 BBSI measurements were available; there were 1182 for the controls. The relationship between visits and these 3367 measures is shown in Table 13-2. The mean atrophy rate calculated for each pairwise combination (i.e. every visit to every other visit) is shown plotted against the mean inter-scan interval for patients (Figure 13-1) and controls (Figure 13-2). The mean rate of atrophy for both groups shows greater variability at short intervals, and less variability at greater intervals. This variability is seen when the standard deviations of these measures are plotted against mean inter-scan follow up (Fig 13-3, patients; Figure 13-4 controls). For both groups the variance (or SD) falls dramatically with increasing interval from 0 to six

152
months; thereafter, slower but detectable reductions in SD continue with increasing intervals through until the end of the study period.

Table 13-2. Number of registrations performed.

For each interval (start scan left column; end scan top row), the number of registrations is shown.

<table>
<thead>
<tr>
<th>Start Scan (Visit - Scan number if 2 performed at visit)</th>
<th>End Scan (Visit - Scan number if 2 performed at visit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>1-2  2  3-1  3-2  4  5-1  6-1  6-2  7  9  10</td>
</tr>
<tr>
<td>1-1</td>
<td>65  66  66  63  67  65  60  57  66  39  22</td>
</tr>
<tr>
<td>1-2</td>
<td>63  63  61  64  62  57  55  63  37  20</td>
</tr>
<tr>
<td>2-1</td>
<td>64  61  65  64  59  56  65  39  22</td>
</tr>
<tr>
<td>3-1</td>
<td>63  66  64  58  55  65  37  20</td>
</tr>
<tr>
<td>3-2</td>
<td>62  61  55  54  61  37  20</td>
</tr>
<tr>
<td>4-1</td>
<td>66  60  57  66  39  22</td>
</tr>
<tr>
<td>5-1</td>
<td>59  56  65  39  22</td>
</tr>
<tr>
<td>6-1</td>
<td>57  60  39  21</td>
</tr>
<tr>
<td>6-2</td>
<td>57  39  21</td>
</tr>
<tr>
<td>7-1</td>
<td>39  22</td>
</tr>
<tr>
<td>9-1</td>
<td>21</td>
</tr>
</tbody>
</table>
Figure 13-1. Mean atrophy rate vs. mean inter-scan follow-up in AD. Atrophy rates (%/yr) are calculated for every possible two-scan combination*

* negative values represent decline in volume (i.e. atrophy)

Figure 13-2. Mean atrophy rate vs. mean inter-scan follow-up in controls. Atrophy rates (%/yr) are calculated for every possible two-scan combination*

* negative values represent decline in volume (i.e. atrophy)
Figure 13–3. Mean standard deviation against mean inter-scan follow-up in AD. SD of atrophy rates (%/yr) are calculated for every possible two-scan combination.

Figure 13–4. Mean standard deviation vs. mean inter-scan follow-up in controls. SD of atrophy rates (%/yr) are calculated for every possible two-scan combination.
13.1.3 Discussion

Using a much larger range of intervals than described in chapter 12, these results demonstrate the relationship between the length of follow-up, mean, and SD of global brain atrophy. Again, these results suggest that, in the absence of an increased rate of drop-out, in a standard placebo controlled trial with two time points, the longest feasible interval is likely to require the smallest sample sizes to adequately power treatment studies.

However, an understanding of the overall "true" mean of these atrophy rates, the "true" variance, and the contribution of inter-subject, intra-subject and measurement error to this variance could allow for more rational decisions regarding trial designs using atrophy based outcome measures. For this to be possible, a multi-level model was developed to determine these parameters from this data-set. Once the mean and variance were established in this way, it was possible to validate the results by comparing the mean and SD predicted by the model with those calculated above.

13.2 Modelling data: theoretical aspects and validation

13.2.1 Introduction

The results presented in this thesis to date have considered a “standard” trial scenario, with atrophy rates calculated from a designated baseline scan over a period of time to a second, follow-up scan. Whilst scenarios such as these are commonly used in clinical trials, the statistical power of longitudinal studies could be increased by making repeated measures (e.g. of atrophy) or more than two time points (Verbeke and Molenberghs, 2000). The advantages of this approach include less reliance on any given scan which may be inaccurate for many reasons, including physiological change, drug changes, scanner shifts. Multiple scans in combination could theoretically reduce "noise" and thus variance from atrophy measurements, thus increasing power to detect change.

The analysis of repeated measures has particular significance when ‘direct’ measures of change are made, as is seen when direct measures of volume loss are automatically calculated using the BBSI. As a BBSI value can be calculated from any pair of
sequential scans, the potential number of measures increases considerably when multiple scanning time points are chosen.

These changes can be illustrated for any one individual (Figure 13-5). The open triangles are measured brain volumes at each of the time points, expressed as a proportion of the initial volume and plotted on a log scale. The solid lines represent the proportional measures of change calculated using the BBSI. To allow comparison between the indirect (brain) and direct (BBSI) measures, the BBSIs have been (arbitrarily) positioned to intersect with the brain volumes at the mid-point of follow-up (Frost et al, 2004). Results are shown over a one-year follow up.

**Figure 13–5.** Multiple direct and indirect measures of atrophy for one individual. Each triangle represents a measure of brain volume. Each line, a BBSI measurement, “hinged” at the midpoint between the two scans to which it refers.

It is apparent from this example that for this individual there is decline in brain volume over time. However, there is a degree of variability between the multiple measures taken, such that any direct measurement (BSI) is associated with an error. These errors, taking any given measurement away from the “true” mean, are likely to be due to a combination of factors, including physiological changes, scanner changes.
and measurement error. Furthermore, one of these graphs with up to 12 measured brain volumes and 66 measured BBSI measurements can be drawn for each of the 69 subjects in this study, each of whom will have a different slope. Figure 13-6 exemplifies the problems that can occur if for one of many reasons (e.g. scanner drift or excessive movement) any given scan or series of scans is associated with an excess degree of error. In this example, all measurements made towards time point 6 (~270 days) are excessively negative, and all from time point 6 are excessively positive. In this case the scans at time point 6 were complicated by excessive movement. Furthermore, although corrected for by the 9dof registration (as seen in the BBSI measurement), the initial scan was affected by significant voxel drift (see chapter 9). It is clear that if a trial depended only on two MRI time points, and the second time point was (in this case) at time point 6, it would clearly be extremely difficult to show a disease modifying effect.

Figure 13–6. Atrophy in an individual: the effect of poor scans.
Each triangle represents a measure of brain volume. Each line, a BBSI measurement, “hinged” at the midpoint between the two scans to which it refers. Significant errors can be seen at the initial visit and visit 6 (~270 days), altering all measurements for this individual away from his “true” slope.
One way to attempt to account for these errors is to model these data mathematically and thus to estimate the true slope (i.e. atrophy) and variance of the slope, and contributors of this variance. With this information, it should then be possible to estimate the sample size implications of any potential trial design involving two or more scanning time points.

One such modelling approach has been designed and validated in this specific scenario (Frost et al, 2004). This approach uses a variant of multi-level modelling (linear mixed models) to develop strategies to distinguish different sources of variability both between and within individuals (Verbeke and Molenberghs, 2000).

The model consists of a number of steps.

1) The basic step is to consider a linear regression model for each individual.

Thus an individual’s brain volume at a given time \( y_{ij} \) is calculated as the expected initial brain volume of that individual \( \alpha_i \), adjusted for the expected decline of that individual \( t_i \) adjusted for the deviation of this individual from his personal mean slope \( b_i \).

\[
y_i = \alpha_i + b_i t_i
\]

2) Multi-level modelling considers each individual’s regression line in terms of the whole group, i.e. adding in a step to describe an individual’s measurement in terms of both slope of that individual, and how that slope differs from that of the whole group.

Thus an individual’s brain volume at a given time \( y_{ij} \) is calculated as the expected initial brain volume of that individual \( \alpha_i \) adjusted for: the expected group decline \( \beta \) adjusted for random deviation of the individual from this mean \( b_i \) over the period of measurement \( t_{ij} \); the underlying within subject deviation from a linear decline in the
absence of measurement error ($u_{ij}$); and the variability introduced when making a measurement ($e_{ij}$).

$$y_{ij} = \alpha + (\beta + b_i) t_{ij} + u_{ij} + e_{ij},$$

[where atrophy rates vary from person to person with mean $\beta; b_i$ is normally distributed with mean 0 and variance $\sigma^2_b; u_{ij}$ is normally distributed with mean 0 and variance $\sigma^2_u; and e_{ij}$ is normally distributed with mean 0 and variance $\sigma^2_e$]

3) In the model designed for the analysis of this data-set, adjustments are made for the fact that direct rather than indirect measurements of change are being made. This has important consequences, as sequential measurements cannot be assumed to be independent. Thus as illustrated in Figure 13-6, a more negatively estimated BSI is more likely to be followed by a more positive BSI

Thus for the direct measured change in brain volume in the $i$th subject between visit $j$ and $k$ ($C_{jk}$):

$$C_{jk} = (\beta + b_i)(t_k - t_j) + u_{ik} - u_{ij} + d_{ijk}$$

where $u_{ik} - u_{ij}$ are the errors associated with making a measurement at respectively the second and first of these time points (e.g. scanning anomalies or segmentation errors). By subtracting the second measure from the first, the expected negative relationship between the two measurements is accounted for. $d_{ijk}$ represents the error in making the direct measurement (i.e. the BBSI), and is normally distributed with mean 0 and variance $\sigma^2_d$.

The features of this model that allow for the expected negative relationship between two consecutive measurements can be illustrated by considering two direct measurements, one between time points 1 and 2; the second between time points 2 and 3:
Time point 1 – 2: \[ C_{i12} = (\beta + b_i)(t_{i2} - t_{i1}) + u_{i2} - u_{i1} + d_{i12} \] (1)

Time point 2 – 3: \[ C_{i23} = (\beta + b_i)(t_{i3} - t_{i2}) + u_{i3} - u_{i2} + d_{i23} \] (2)

It can be seen that the common shared term \( u_{i2} \), which is positive in the first and negative second, provides the term allowing for the expected negative correlation. Thus going from time point 1 to time point 3 in two steps, entails adding (1) and (2) together, giving:

\[ C_{i12} + C_{i23} = (\beta + b_i)(t_{i2} - t_{i1}) + u_{i2} - u_{i1} + d_{i12} + (\beta + b_i)(t_{i3} - t_{i2}) + u_{i3} - u_{i2} + d_{i23} \]

(3)

By contrast, one direct measure from scan 1 to scan 3, is equal to:

\[ C_{i13} = (\beta + b_i)(t_{i3} - t_{i1}) + u_{i3} - u_{i1} + d_{i13} \] (4)

Comparing (3) and (4), it can be seen that one direct measure of change (4) differs from two additive measures of change (3) merely in terms of the direct measurement errors, i.e. \( d_{i13} \) vs. \( (d_{i12} + d_{i23}) \)

4) In this model, an additional adjustment is made to account for the fact that two measurements on the same day are likely to be associated with different, (probably) smaller variability than two sequential BBSIs taken on different days. Thus whilst there is likely to be small variability in measures of brain volume between two sequentially acquired scans (e.g. due to movement, subtle moment-to-moment changes in scanner gradients), these are likely to be less than changes associated with two scans acquired from the same individuals over a longer time period (e.g. due to more major scanner fluctuations or larger physiological changes including hydration).
The model thus described can be used to analyze any series of consecutive BBSI data, in statistical packages such as SAS. The model produces estimates of the following parameters for each group (i.e. patients or controls):

Mean slope \((m)\): i.e. mean atrophy rate of the group

Variance between individuals \((\sigma_b^2)\): i.e. the variance of \(b_i\)

Variance both between visits and between scans (together: \(\sigma_u^2\)): i.e. the variance of \(u_i\)

Variance associated with making a direct measurement (i.e. measuring BBSI) \((\sigma_d^2)\)

Thus it is possible to determine the mean slope \((m)\), inter-individual variation \((\sigma_b^2)\) and intra-individual variation \((\sigma_u^2 + \sigma_d^2)\), which can then be used to estimate sample size calculations, and determine the effect of any possible design of trial strategy on possible power calculations. The aim of this modelling approach is to avoid the dangers of an ad-hoc approach - simply trying one analysis strategy after another - which as demonstrated leads to highly variable results. Details of the full model have previously been validated on a small data set (12 patients with AD) (Frost et al, 2004).

13.2.2 Methods

In order to determine the above parameters, the entire data set of BBSI measurements was modelled separately for both patients and controls. To allow comparison between direct and indirect BBSIs, results were analyzed on a log scale. Data were prepared in Stata Version 8, and modelling was performed in SAS Version 8.2. Results for \(m\), \(\sigma_u^2\), \(\sigma_d^2\) and \(\sigma_b^2\) were obtained. These estimates were then used to determine the "true" mean and variance for BBSI derived atrophy rates over a number of intervals (excluding technical factors including missed scans, unusable scans or drop-outs). As inter-individual variance changes proportional to the square of the time interval, over one year, the expected variance is equal to:

\[
\text{Variance (SD}^2) = \sigma_b^2 + (2\sigma_u^2 + \sigma_d^2)
\]

To validate the model further, using the derived mean and variance parameters, the estimated mean atrophy rate for the patients was plotted against the directly measured
results, and the estimated SD was plotted against the previously determined SD as it changed with interval.

13.2.3 Results

<table>
<thead>
<tr>
<th></th>
<th>Alzheimer’s disease</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean slope (m)</td>
<td>2.23</td>
<td>0.49</td>
</tr>
<tr>
<td>Between subject variance ($\sigma_b^2$):</td>
<td>0.97</td>
<td>0.0</td>
</tr>
<tr>
<td>Between visit variance +</td>
<td>0.22</td>
<td>0.16</td>
</tr>
<tr>
<td>Between scan variance</td>
<td>0.12</td>
<td>0.02</td>
</tr>
<tr>
<td>=Within subject variance ($\sigma_u^2$)</td>
<td>0.34</td>
<td>0.18</td>
</tr>
<tr>
<td>Variance of BBSI measurement ($\sigma_d^2$)</td>
<td>0.0025</td>
<td>0.0004</td>
</tr>
</tbody>
</table>

Thus for the estimated true mean and SD atrophy rate at one year was:

$2.23 \pm \sqrt{\sigma_b^2 + 2\sigma_u^2 + \sigma_d^2}$, or $2.23 \pm 1.28\%$ per year for patients

$0.49 \pm \sqrt{\sigma_b^2 + 2\sigma_u^2 + \sigma_d^2}$, or $0.49 \pm 0.6\%$ per year for controls

For the patients, the predicted mean atrophy rate was plotted against the observed mean atrophy rate, as previously determined (Figure 13-7); the predicted standard deviation was plotted against the observed standard deviation (Figure 13-8). Figure 13-7 shows that the predicted mean reflects the observed values, which are clustered symmetrically around it, and at longer intervals (over which the measurements are more accurate), closely approximates to it. Figure 13.8 shows an extremely close approximation of the estimated to observed SDs.
Figure 13-7. Observed mean atrophy rates vs. model-derived mean (line).

Figure 13-8. Observed SDs atrophy rates vs. model-derived SDs (line).
13.2.4 Discussion

This work confirms that the statistical approach designed to model data from longitudinal studies using direct measures of change (e.g. the BBSI) is able to estimate the observed means and standard deviations in AD and controls with a high degree of accuracy. This approach can be used to analyze any serial MRI data-set using the BBSI, or other direct measures of change, as an atrophy measure. The derived results accurately define the mean, the variance, and the contribution of error terms to the variance.

The mean atrophy rate in cases was determined to be 2.23% per annum, a rate very similar to previously reported rates of whole brain atrophy which have ranged from 2.0 – 2.4% per annum (Fox et al, 1999a; Fox et al, 2000; O'Brien et al, 2001; Bradley et al, 2002; Wang et al, 2002). The standard deviation at one year was estimated to be 1.28% per annum, again comparable with previous results of 0.9-1.4% (Fox et al, 1999a; Fox et al, 2000; O'Brien et al, 2001; Wang et al, 2002). Only Bradley et al found a substantially smaller SD (0.5), based on a mixed linear model of 5 subjects (Bradley et al, 2002). It is likely that the results from the work presented reflect the mean and standard deviation of an elderly cohort of patients with a range of disease presentation and severities. When the components of variance were assessed, it was estimated that the between subject variance was approximately three times the within subject variance at one year. This suggests that compared to the group as a whole, an individual with AD undergoes atrophy at a relatively stable rate; however both these errors are of a different order of magnitude to that associated with making a direct measure of atrophy (i.e. the BBSI) in which the variance (and hence SD) was found to be negligible (0.0025%).

The rate of atrophy associated with normal ageing was found to be 0.49% per annum, in line with several reports finding atrophy rates of 0.3-0.5% per annum in healthy adults in the 70s and 80s (Resnick et al, 2003; Mueller et al, 1998; Scahill et al, 2003; Fox et al, 1999a). It is interesting to note that the results of this analysis suggest that there is effectively no inter-individual variation in rates of atrophy between normal subjects, and that nearly all sources of variance come from within-subject variability, with a minimal contribution coming from the (very robust) direct BBSI measure. This
would suggest that all elderly controls appear to lose brain volume at the same (or at least very similar) rates. Within subject variability was slightly smaller in controls than patients. This is likely to be related to a number of factors including more movement, and more inherent variability (i.e. slow acceleration of atrophy over the study (Chan et al, 2003)) in the AD group. Nonetheless, in a clinical trial, if allowances are to be made for the effect of normal ageing (i.e. the assumption that the best a drug could do would be to reduce the rate of atrophy to that of normal ageing), it seems reasonable to assume that with this technique, healthy control subjects in this cohort aged ~ 70 years lose brain volume at a rate of 0.49% per annum.

With the results calculated from this analysis, it was then possible to assess a variety of trial designs, and the estimated sample sizes required to power them. In addition, the model provides a framework to explore the determinants of inter- and intra-individual variability. These are the subjects of the following two chapters.
14 Trial design using modelling results

14.1 Introduction

With an accurate understanding of the components of variance, it is possible to estimate the sample sizes that would be required not only for those intervals studied, but for many other possible trial designs. This allows for trial designs with more than two MRI trial points to be assessed. As discussed previously, it is possible that by combining more than one interval sources of error could be reduced, in turn reducing the number of subjects required.

14.1.1 Statistical background

14.1.1.1 Two-time point designs

As discussed previously, statistical power is principally determined by the variance of the measurement in question. For a simple two-time point trial design, the variance can be estimated from the components estimated previously, to be:

\[ \sigma_b^2 + (2\sigma_u^2 + \sigma_d^2)/x^2 \]

where \( x \) is the number of years between the two scans.

Thus, at one year:

\[ \sigma_b^2 + (2\sigma_u^2 + \sigma_d^2) \]

Whilst at six months:

\[ \sigma_b^2 + (2\sigma_u^2 + \sigma_d^2)/0.5^2 \]
\[ = \sigma_b^2 + 8\sigma_u^2 + 4\sigma_d^2 \]

14.1.1.2 Multiple time-point designs

The multi-level model design used to estimate sources of variance (chapter 13), can be extended to assess the effect of variance over multiple time-points. This requires a calculation of the relative weight that each scan time-point should contribute to the variance estimate.
In the model defined for the analysis of this data-set (Frost et al, 2004), a general equation involving a series of matrices is defined to describe the contribution of multiple measurements to the overall slope. Thus, for a four time-point design, six direct measures can be made e.g. time point one to two ($C_{12}$), one to three ($C_{13}$), one to four ($C_{14}$), two to three ($C_{23}$), two to four ($C_{24}$), and three to four ($C_{34}$). These can be illustrated using the following formula, where the design matrix is used to determine which measurement ($C_i$) is associated with which measures of time ($t_i$), intra-subject error ($u_i$) and direct measurement error ($d_i$).

\[
C_i = N_t (\beta + b_i) + N_u u_i + d_i
\]

Where $C_i$ is the vector for measures of direct change for the $i$th subject, $t_i$ is the vector of measurement times, and $u_i$ and $d_i$ are the vectors of within-subject and direct measurement random error respectively. $N$ is the design matrix determining which are the start (-1) and end visits (1) for each measure of $C_i$. A feature of multi-level models is that the variance from a number of measurements can be calculated to take into account the correct weighting of each interval, according to the formula:

\[
\text{Variance} = (X^T V^{-2} X)^{-1}
\] (Goldstein, 1995)

\[1\]
Where $X$ and $X^T$ are matrices converting the variance ($V$) matrix back to numerical form. The variance ($V$) in this case is given by the equation:

$$
\sigma_b^2 N t^t N + \sigma_u^2 NN^t + \sigma_d^2 I
$$

(2)

Where $N_t$ is the vector combining the design matrix multiplied by the time matrix; $t^tN^t$ is the transposed matrix $N_t$ i.e. converted from an (x by 1 matrix) to a (1 by x matrix); $NN^t$ is the design matrix multiplied by an transposed design matrix; and $I$ is the identity matrix (i.e. a diagonal matrix with all diagonal element = 1)

By combining equations (1) and (2), it is possible to determine a formula to describe the optimal estimate of the atrophy rate for an individual over any series of time intervals. Knowing the mean atrophy rate of the group (2.23), and the various sources of variance ($\sigma_b^2 = 0.97; \sigma_u^2 = 0.34; \sigma_d^2 = 0.0025$), it is possible to determine the true variance of any serial scan design; this in turn allows for calculations of the numbers needed to power a study with any given trial design.

14.2 Methods

Using the equations given above and the mean and sources of variance derived in chapter 8, calculations were performed to estimate sample sizes that would be required to power placebo-controlled studies over six months, one year and two years.

Calculations were first performed to determine the sample size required in each arm of a placebo controlled study to detect 20% reduction in atrophy (i.e. from 2.23 to 1.78 % per annum). Under these assumptions, the maximum effect that a drug could produce would be to reduce atrophy from 2.23% per annum to zero. As normal controls lose atrophy at a rate of 0.49% per annum, it is perhaps more plausible that the maximum effect of a drug might be to reduce rate of atrophy from 2.23% to 0.49%. In order to estimate numbers needed to produce a true 20% reduction in excess disease-related atrophy, sample size calculations were also performed to reduce atrophy rate from 2.23% to 1.88% (i.e. 20% of 2.23-0.49).
Power calculations were performed with a number of design strategies, ranging from two-scan time points to seven-scan time points. The effect of adding in additional scans, and altering the timing of these additions was assessed. All calculations were performed using standard formulae (Armitage and Berry, 1991), with 90% power to detect change at a 5% (two tailed) significance level.

14.3 Results

Results of sample size estimates for a variety of trial designs over a one-year periods are shown in Table 14-1. Sample sizes can be reduced by approximately 20% by increasing the number of scans obtained from two to seven. It is notable that scans clustered around the start or end of the study improve sample sizes the most, by providing longer periods over which the measurements can contribute.

Over two years, relatively small benefits can be gained by increasing the number of scans acquired (Table 14-2). Thus going from two scans separated by a 24 month interval, to seven scans acquired over the period decreases the sample size by only about 9%.

The biggest gains can be seen over a six month period (Table 14-3). By increasing the number of scans to 7 and ensuring optimal spacing (i.e. at the beginning and end of follow-up), it is possible to reduce sample sizes from 390 to 233 per treatment arm, a reduction of 43%, and a real saving of 314 patients in such a trial.

For all of these trial designs, it can be seen that it is not only the number but also the timing of the scans that is important. Thus for three time point designs, no difference in the estimated sample size is seen if a third scan is placed exactly half way between the first and last scans. However considerable reductions in sample sizes may be possible if the additional scan visit is moved either close to the first or last scan visit (Table 14-3).
Table 14-1. Estimated sample sizes to power each arm of a one year study. Results with a number of different scenarios are shown, with and without correction for normal ageing.

<table>
<thead>
<tr>
<th>Number of scans</th>
<th>Timing of scans (months)</th>
<th>Sample size*</th>
<th>Sample size corrected for normal ageing**</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0,12</td>
<td>177</td>
<td>291</td>
</tr>
<tr>
<td>3</td>
<td>0,6,12</td>
<td>177</td>
<td>291</td>
</tr>
<tr>
<td>3</td>
<td>0,1,12</td>
<td>163</td>
<td>268</td>
</tr>
<tr>
<td>4</td>
<td>0,4,8,12</td>
<td>170</td>
<td>279</td>
</tr>
<tr>
<td>4</td>
<td>0,1,11,12</td>
<td>148</td>
<td>244</td>
</tr>
<tr>
<td>5</td>
<td>0,1,6,11,12</td>
<td>148</td>
<td>244</td>
</tr>
<tr>
<td>5</td>
<td>0,1,2,11,12</td>
<td>145</td>
<td>237</td>
</tr>
<tr>
<td>6</td>
<td>0,1,2,10,11,12</td>
<td>140</td>
<td>229</td>
</tr>
<tr>
<td>7</td>
<td>0,1,2,6,10,11,12</td>
<td>140</td>
<td>229</td>
</tr>
<tr>
<td>7</td>
<td>0,1,2,3,10,11,12</td>
<td>138</td>
<td>227</td>
</tr>
</tbody>
</table>

*: 20% reduction in total atrophy (i.e. reducing rate from 2.23% to 1.78% per year)

**: 20% reduction in disease related atrophy (i.e. reducing rate from 2.23% to 1.88% per year).
Table 14-2. Estimated sample sizes to power each arm of a two year study. Results with a number of different scenarios are shown, with and without correction for normal ageing.

<table>
<thead>
<tr>
<th>Number of scans</th>
<th>Timing of scans (months)</th>
<th>Sample size*</th>
<th>Sample size corrected for normal ageing**</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0,24</td>
<td>124</td>
<td>204</td>
</tr>
<tr>
<td>3</td>
<td>0,12,24</td>
<td>124</td>
<td>204</td>
</tr>
<tr>
<td>3</td>
<td>0,2,24</td>
<td>121</td>
<td>198</td>
</tr>
<tr>
<td>4</td>
<td>0,8,16,24</td>
<td>122</td>
<td>200</td>
</tr>
<tr>
<td>4</td>
<td>0,2,22,24</td>
<td>117</td>
<td>192</td>
</tr>
<tr>
<td>5</td>
<td>0,2,12,22,24</td>
<td>117</td>
<td>192</td>
</tr>
<tr>
<td>5</td>
<td>0,2,4,22,24</td>
<td>116</td>
<td>190</td>
</tr>
<tr>
<td>6</td>
<td>0,2,4,20,22,24</td>
<td>114</td>
<td>188</td>
</tr>
<tr>
<td>7</td>
<td>0,2,4,12,20,22,24</td>
<td>114</td>
<td>188</td>
</tr>
<tr>
<td>7</td>
<td>0,2,4,6,20,22,24</td>
<td>114</td>
<td>188</td>
</tr>
</tbody>
</table>

*: 20% reduction in total atrophy (i.e. reducing rate from 2.23% to 1.78% per year)

**: 20% reduction in disease related atrophy (i.e. reducing rate from 2.23% to 1.88% per year).
Table 14-3. Estimated sample sizes to power each arm of a six month study. Results with a number of different scenarios are shown, with and without correction for normal ageing.

<table>
<thead>
<tr>
<th>Number of scans</th>
<th>Timing of scans (months)</th>
<th>Sample size *</th>
<th>Sample size corrected for normal ageing **</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0,6</td>
<td>390</td>
<td>640</td>
</tr>
<tr>
<td>3</td>
<td>0,3,6</td>
<td>390</td>
<td>640</td>
</tr>
<tr>
<td>3</td>
<td>0,0.5,6</td>
<td>336</td>
<td>551</td>
</tr>
<tr>
<td>4</td>
<td>0,2,4,6</td>
<td>361</td>
<td>592</td>
</tr>
<tr>
<td>4</td>
<td>0,0.5,5,5.5,6</td>
<td>273</td>
<td>448</td>
</tr>
<tr>
<td>5</td>
<td>0,0.5,3,5.5,6</td>
<td>273</td>
<td>448</td>
</tr>
<tr>
<td>5</td>
<td>0,0.5,1,5.5,6</td>
<td>258</td>
<td>423</td>
</tr>
<tr>
<td>6</td>
<td>0,0.5,1,5,5.5,6</td>
<td>239</td>
<td>392</td>
</tr>
<tr>
<td>7</td>
<td>0,0.5,1,3,5.5,5.5,6</td>
<td>239</td>
<td>392</td>
</tr>
<tr>
<td>7</td>
<td>0,0.5,1,1.5,5,5.5,6</td>
<td>233</td>
<td>382</td>
</tr>
</tbody>
</table>

*: 20% reduction in total atrophy (i.e. reducing rate from 2.23% to 1.78% per year)

**: 20% reduction in disease related atrophy (i.e. reducing rate from 2.23% to 1.88% per year).
14.4 Discussion

These results demonstrate that, particularly over short intervals, samples sizes may be considerably reduced if more than two imaging time points are used. Over a 24 month period, relatively little gain can be made by adding extra scans, principally because over long periods, the variance of the mean atrophy rate (the variance decreases by the square of the interval) is small and stable. By contrast, over a six month interval, sample sizes can be considerably reduced by adding extra scans. However, to have maximum effect, these extra scans need to be added either close to the end or the beginning of the study; thus at six months, adding an extra three month scan to increase a six study trial design to a seven scan trial, makes no material difference to the estimated sample sizes. As discussed below, in order to maximize information from patients dropping out during the study, it is likely to be most practical to place additional scans at the end rather than the beginning of any study design.

There are additional advantages and some disadvantages that may come from study designs utilizing multiple scanning time points. Whilst additional scans impose extra expenses on a study, these costs are likely to be outweighed by the savings in reducing patient numbers. This approach may require more frequent visits over a shorter time frame; such an intense schedule, although onerous to the participants, may actually improve retention in a study, and by reducing the time frame over which a study may be run, it may be possible to reduce drop-outs (see chapter 6). There may be additional advantages in multiple scan point trials in terms of gaining information from patients who drop-out of a study prematurely. If a traditional two time-point trial were to be run over a six month period, a patient dropping out at 5½ months would provide no atrophy data for analysis. If however, the study design had included scans at 0, 5 and 6 months, an atrophy measurement for this individual could be calculated from the 0 - 5 month interval, and could be used in an intention-to-treat analysis.

An additional benefit of running studies with multiple scanning points may be that the problems associated with isolated poor quality scans may be reduced; this is particularly the case when a multi-level model approach, which can compensate for the "kicks" associated with bad scans, is used.
The sample size estimate at one year using the optimal combination of seven scans was 138 patients per treatment arm, rising to 227 if the effect of normal ageing was taken into account. These sample sizes are based on a large prospectively studied cohort, and compare favourably with BBSI derived estimates previously reported using smaller numbers of often younger patients (Fox et al, 1999a; Fox et al, 2000; O'Brien et al, 2001). The six month estimates, previously far higher than those calculated using ventricular measures, are now much more comparable (233 with the BBSI versus 170-200 with ventricular measures). Thus the BBSI, a direct measure of brain atrophy, is also a feasible outcome measure in six-month studies, particularly if multiple scan time-points are acquired. It may be that a combination of ventricular measures (providing increased power) combined with BBSI derived whole brain atrophy rates (which are likely to prove more biologically plausible surrogates of AD progression and more robust to confounds such as hydrocephalus) could overall improve the design of such short interval studies. Such studies are likely to be strongly influenced by the timing of the scans which on the basis of these results should be weighted towards the start and end of the study.
15 Determining correlates of increased atrophy

15.1 Introduction

The variability in atrophy rate between individuals demonstrated using the modelling approaches provides an opportunity to compare rates of decline in atrophy with other measures of longitudinal change. Decline in both function and neuropsychological performance is accepted as inevitable features of AD, and mortality in AD has been found to be strongly associated with rate of cognitive decline (Hui et al, 2003). It follows that demonstration of correlation between decline in these measures and decline in atrophy could provide further evidence to strengthen the case for global brain atrophy as a valid surrogate marker of AD progression (see chapter 3). Few studies to date have directly compared progressive atrophy with change in longitudinal measures. In one such study, Fox et al demonstrated a close relationship between increasing atrophy and decline in MMSE score over an interval of up to five years (Fox et al, 1999a). More recently Jack et al (Jack et al, 2004) demonstrated a correlation between both whole-brain and ventricular change and three measure of cognition in 64 patients with AD studied over intervals of ~1.3 years. Despite the relatively small absolute sample sizes in this study, the large number of individual measures made per subject, and the multi-level model designed to describe them, provide a framework in which potential longitudinal correlates of atrophy may be tested.
15.2 Methods

To determine statistically significant longitudinal correlates of atrophy, the annualized percentage change of each of the following parameters was calculated for each patient: RMT-words grade-score change; RMT-faces grade-score change; composite memory z-score change; silhouettes z-score change; arithmetic z-score change; naming z-score change; vocabulary z-score change; matrix reasoning z-score change; composite memory z-score change (see chapter 6). The multi-level model previously described was extended to determine the dependency of atrophy on correlates for the AD group. Modelling was performed in SAS Version 8.2. Longitudinal rates of change were then entered into the model (for the patients with AD) as potential covariates of atrophy; significance was determined at the $p<0.05$ level.

15.3 Results

Significant positive correlation with global brain atrophy rate was seen in 3 of the 9 potential correlates tested; these are shown in Table 15-1. Change in vocabulary z-score just failed to reach significance ($p=0.06$).

Table 15-1. Statistically significant longitudinal correlates of atrophy

<table>
<thead>
<tr>
<th>% change in naming z-score</th>
<th>1693</th>
<th>-2.02</th>
<th>-2.558</th>
<th>-3.096</th>
<th>0.022</th>
</tr>
</thead>
<tbody>
<tr>
<td>% change in silhouettes z-score</td>
<td>1783</td>
<td>-2.13</td>
<td>-2.634</td>
<td>-3.138</td>
<td>0.031</td>
</tr>
<tr>
<td>% change composite cognitive z-score (excluding memory)</td>
<td>2029</td>
<td>-1.84</td>
<td>-3.12</td>
<td>-4.4</td>
<td>0.006</td>
</tr>
</tbody>
</table>
15.4 Discussion

These results suggest that change in tests of naming, silhouettes or a global cognitive score (excluding memory) significantly correlated with increased rates of atrophy. Change in vocabulary z-score just failed to reach significance ($p=0.064$); none of the other scores (including those sensitive to memory) or change in the MMSE showed significant correlation with atrophy rates.

The finding of significant correlation between change in several cognitive scores and atrophy provides further supportive evidence for atrophy as a surrogate marker of AD progression. It is interesting to note that changes in memory scores did not correlate with atrophy; as memory is an early and prominent feature of AD, it is likely that despite the use of an easy memory test with a wide dynamic range, severely affected patients may reach floor, or near floor levels, whilst more mildly affected patients continue to decline. This is supported by the finding that in this study, at the first assessment 33% of patients were already scoring at floor on the RMT-words test, and 40% on the RMT-faces. By the end of the study, these figures were 58% and 53% respectively. It is perhaps not unexpected that there is no significant relationship between decline in memory and atrophy, as the latter continues to decline throughout the disease process (Silbert et al, 2003). This is further supported by the results of Jack et al (Jack et al, 2004) who did not find a consistent relationship between whole brain (nor medial temporal atrophy) and decline in memory.

Cross-sectional studies have demonstrated relationships between cognitive decline in AD and MRI measures of brain (Murphy et al, 1993) and hippocampal volume (Jack et al, 1997), and cortical atrophy at post-mortem (Mouton et al, 1998). However, only longitudinal studies can determine correlations between change in cognitive measures and longitudinal measures of atrophy. In one such longitudinal study, Fox et al demonstrated a close correlation between rate of global cerebral atrophy and change in MMSE in 29 patients with AD assessed over periods of up to 6 years (Fox et al, 1999a). It is notable that from the results described in this thesis, whilst neuropsychological measures showed correlation with atrophy rate, the same was not true for MMSE change. This discrepancy is likely to be due to the inherent variability...
of the MMSE as a progression marker over short intervals (Clark et al, 1999; Galasko et al, 2000), complicated by the observed interaction with AchEI (chapter 6).

15.5 Conclusions

The results of this work suggest that change in a number of non-memory neuropsychological scores is significantly correlated with increased rates of cerebral atrophy. This provides further evidence to suggest that cerebral atrophy is a valid biomarker of AD progression, and that atrophy may be a useful marker of disease progression in clinical trials, possibly in conjunction with standard clinical and neuropsychological measures. However, ultimate validation of imaging as a surrogate marker of therapeutic efficacy rather than as a marker of disease progression in a natural history study such as this, requires a positive disease-modifying therapeutic trial, which to date has yet to occur (Jack et al, 2004; Fox et al, 1999a).
16 Determining predictors of increased atrophy rates

16.1 Introduction

The results from the multi-level model (described in chapter 13) indicate that for this cohort of patients, the majority (approximately three-quarters) of variability in rate of atrophy in AD is due to variation between-subjects. This inter-subject variation is also seen anecdotally in clinical practice: whilst some patients with AD run a relatively slow course, others decline much more rapidly. An understanding of which factors are predictive of subsequently increased rates of brain atrophy might therefore be useful in providing prognostic information to individual patients. Furthermore, this information might be useful in a trial setting, potentially allowing for studies in more homogeneous groups with smaller inter-individual variability, thus reducing sample size requirements; alternatively an understanding of these factors could allow stratification (or other) approaches to the analysis. In order to explore the factors that determine subsequently altered rates of cerebral atrophy, traditional epidemiological studies involve following large cohorts of patients over long periods time. Whilst the size of this cohort is too small and the duration of follow-up too short to undertake these types of analysis, the large number of individual measures made per subject, and the multi-level model designed to describe them, provide a framework in which exploratory studies of potential predictors of atrophy may be performed.

16.2 Methods

In order to determine potential cross-sectional (baseline) predictors of increased atrophy rates, the multi-level model previously described was extended to determine the dependency of atrophy rate on covariates for the AD group. The following potential covariates were each assessed: gender; handedness; ApoE4 (one or two alleles vs. none); age at study start; onset age; years since symptom onset; start MMSE; start CDR; CDR sum of boxes; school leaving age; smoker (ever vs. never); systolic blood pressure; diastolic blood pressure; mean blood pressure; BMI; modified Hachinski score; baseline RMT-words grade-score; baseline RMT-faces grade-score; baseline composite memory z-score; baseline silhouettes z-score; baseline arithmetic z-score; baseline naming z-score; baseline vocabulary z-score; baseline matrix reasoning z-score; and baseline composite cognitive z-score (see chapter 5).
Modelling was performed in SAS Version 8.2. Each potential covariate was entered into the model independently, and assessed for significance ($p<0.05$). Covariates found to influence rates of atrophy significantly were then assessed in a forward stepwise model to determine statistically significant independent predictors of increased atrophy.

16.3 Results

Of the 25 potential predictors of atrophy assessed, when assessed individually, six baseline measures were found to be significant predictors of subsequent increased atrophy. These were: younger age at onset; younger age at the start of the study; lower systolic blood pressure; and lower baseline score on arithmetic, matrix reasoning and composite memory (Table 16-1). Lower baseline MMSE had a directional effect to increase rate of atrophy that just failed to reach significance ($p=0.059$). Whilst there was a directional effect of ApoE4 genotype reducing atrophy rate in a dose-dependent manner, this was not significant ($p=0.52$).

The most significant effect (lowest $p$ value) was the baseline matrix reasoning z-score. Using a forward stepwise regression model to account for covariates, baseline matrix reasoning z-score was the only statistically significant independent predictor of increased atrophy. The matrix reasoning results suggest that global brain atrophy accelerated as the severity of the disease increases. With a matrix reasoning z-score of 0 (50th percentile), the expected rate of atrophy over the following year would be: 1.87%/yr; with a z-score of -1 (34th percentile) this would increase to 2.2% per year. If a patient had a matrix reasoning z-score of -2 at the start of the study (5th percentile) the predicted atrophy rate over the following year would be 2.53%/yr. Including matrix reasoning z-score into the multi-level model gave the following estimates of the contribution of variance: $\sigma^2_b$ (between subject variance) = 0.84; $\sigma^2_u$ (within subject variance) = 0.28; $\sigma^2_d$ (BBSI measurement error) = 0.0023. At one year, including matrix reasoning z-score as a covariate, the inter-individual SD was reduced by ~8% (from 0.99 to 0.91%/yr), and the total SD by ~7% (from 1.28 to 1.19). In a six-month seven time-point trial with optimum scan weighting (see Table 14-3), this represents a marginal (~10%) reduction in required sample size from 233 to 213 per treatment arm.
**Table 16-1. Statistically significant baseline predictors of atrophy**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Observations</th>
<th>Atrophy rate (%/yr) at group mean*</th>
<th>Change (%/yr) with each additional unit†</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at study start</td>
<td>2199</td>
<td>2.25</td>
<td>-0.052</td>
<td>0.022</td>
</tr>
<tr>
<td>Age at disease onset</td>
<td>2199</td>
<td>2.22</td>
<td>-0.056</td>
<td>0.013</td>
</tr>
<tr>
<td>Systolic blood pressure</td>
<td>2095</td>
<td>2.29</td>
<td>-0.020</td>
<td>0.030</td>
</tr>
<tr>
<td>Arithmetic z-score</td>
<td>1937</td>
<td>2.33</td>
<td>-0.32</td>
<td>0.026</td>
</tr>
<tr>
<td>Matrix reasoning z-score**</td>
<td>2056</td>
<td>2.26</td>
<td>-0.33</td>
<td>0.004</td>
</tr>
<tr>
<td>Composite cognitive z-score</td>
<td>2112</td>
<td>2.28</td>
<td>-0.35</td>
<td>0.013</td>
</tr>
</tbody>
</table>

*: i.e. age at study start: 69.4 years; age at disease onset: 64.7 years; systolic BP: 143mmHg; arithmetic z-score: -2.03; matrix reasoning z-score: -1.19; composite cognitive score: -1.7.

†: i.e. each additional year at study or disease onset; each additional mmHg; each additional one-point z-score

**: only statistically significant independent predictor of increased atrophy

### 16.4 Discussion

These results suggest a number of baseline factors that may predict subsequently increased rates of cerebral atrophy in AD. In this study, baseline performance on a variety of neuropsychological tests (arithmetic, fluid intelligence, and a composite cognitive score) were all associated with subsequently increased rates of global atrophy; however in the multivariate analysis, matrix reasoning was the only statistically significant independent predictor of increased atrophy. It is notable that no memory score was predictive of decline; this is likely to be due to the fact that
many patients were already performing at or near floor levels on these tests at the start of the study; thus memory scores were not able to distinguish between patients impaired beyond a certain level. Several previous studies have demonstrated that low MMSE score is predictive of subsequent functional decline (O'Hara et al, 2002; Santillan et al, 2003; Doody et al, 2001). In this study, MMSE directionally predicted subsequently increased rates of atrophy, but just failed to reach significance. This is likely to be due to the relatively small numbers of patients in this sample, as well as the limitation of the MMSE in terms of ceiling and floor effects, and reproducibility (see chapter 5). In this setting and in this patient group, this study provides some evidence to suggest that neuropsychological tests of arithmetic, composite (non-memory) cognition and particularly matrix reasoning may be more accurate means of assessing cross-sectional disease severity than the MMSE.

The results from this study suggest that in a univariate analysis, young age at onset (and hence young age at study entry) was associated with increased rates of cerebral atrophy. Several studies, using both neuropsychometric scores (O'Hara et al, 2002; Jacobs et al, 1994) and cross-sectional MR imaging approaches (Woo et al, 1997), have also suggested that young age at onset may predict increased future decline, although other studies have not found this association (Burns et al, 1991). Taken together, these findings suggest that patients developing the disease at younger ages may be more likely to run a rapid disease course. However, in the multivariate analysis, age at onset had no significant effect once severity (as assessed by matrix reasoning) was taken into account, suggesting that the effect of early disease onset on atrophy is in fact mediated via increased disease severity. This is borne out by further analysis of the data: for patients less than 70 at the start of the study the mean matrix reasoning z-score was -1.92 (n=23); for those over 70 the mean was -0.31 (n=19). There are a number of reasons why younger patients in this study may have been more likely to have increased disease severity than their older counterparts. Patients who are older with more severe disease may be less likely to be referred to a specialist neurology cognitive clinic, and more likely to be managed by local services. Furthermore the spouses or carers of older, more cognitively affected patients are themselves likely to be elderly, and may be less willing to be involved in trials, particularly if they are as onerous as this one. By contrast patients with younger age at
onset and their spouses may be both more motivated and able to attend for research studies even when the disease has progressed to a moderately severe stage.

The third predictor of subsequently increased rates of atrophy in this study was low systolic blood pressure. Whilst, as previously discussed, midlife hypertension is considered to be a risk factor for the development of AD (Skoog and Gustafson, 2003), by the time AD is manifest, blood pressure conversely tends to be lower than in age-matched controls and appears to decrease with dementia severity (Davis et al, 2003; Skoog et al, 1996); furthermore lower systolic blood pressure predicts poorer performance on neuropsychological testing (Davis et al, 2003). Whilst some authors have speculated that systolic hypotension may be either a cause or consequence of the progression of AD, in a cross-sectional CT study, lower systolic blood pressure was correlated with increased cross-sectional frontal, parietal and cortical atrophy in both aged controls and patients with AD (Skoog et al, 1998); these authors concluded that hypotension in dementia was likely to be a secondary phenomenon. The results described in this thesis provide some support for this hypothesis: low systolic blood pressure no longer had an effect on subsequent rate of atrophy once performance on matrix reasoning was accounted for, suggesting that thus systolic hypotension may be a marker of disease severity, which in turn is the true predictor of increased rates of atrophy. This hypothesis is supported by the finding that there was a trend for patients with lower systolic blood pressure (<140mmHg) to perform worse on the matrix reasoning z-score (-1.59, n=14) than those with higher blood pressure (≥140mmHg) who had an average score of -0.99 (n=28).

The matrix reasoning subsection of the WASI comprises four types of nonverbal reasoning tasks, including: pattern completion, classification, analogy and serial reasoning, and involves the subject examining a matrix from which a section is missing and determining the means of completing the matrix by selecting one of five potential options (Wechsler, 1999). Matrix reasoning is a predominantly non-verbal score of "fluid intelligence" and reasoning. The test takes approximately ten minutes to perform, and has a wide dynamic range, ensuring that both normal controls and patients with moderately advanced AD can be tested. It is interesting to note that in a prospective study of presymptomatic individuals at risk for familial AD, Fox et al reported that aside from memory, performance IQ (a component of which is fluid
intelligence) was the only other cognitive measure which predicted which individuals proceeded to develop AD (Fox et al, 1998). Taken together, these results suggest that measures of fluid intelligence (and hence performance IQ) may be useful measures to predict future decline in AD.

Whilst there was a directional trend for patients with ApoE4 alleles to have a reduced rate of global brain atrophy, this was not significant \((p=0.54)\). This result may reflect that ApoE4 does not substantially influence rates of whole brain atrophy, or that this study with small numbers of subjects, the majority of whom were ApoE4 heterozygotes, did not have sufficient power to determine such an effect. The ApoE4 allele is known to reduce the age at onset of the disease (Higgins et al, 1997), and has been reported to increase the rate of hippocampal atrophy in a dose dependent manner (Mori et al, 2002). It is therefore possible that the presence of an ApoE4 allele could affect relative regional rates of atrophy (leading to phenotypic differences) without significantly altering rates of whole brain atrophy; alternatively, an effect of ApoE on global brain atrophy would be shown in a sufficiently large study.

Results produced from the multi-level model suggest that atrophy rate in AD increases as the disease becomes more severe. Thus every one point fall in matrix z-score is associated with a 0.33% increase in annual atrophy rate. Extrapolating backwards, to a time when the patients’ mean matrix z-score was the same as the controls (i.e. 0.98 instead of -1.19), assuming the relationship between atrophy and matrix z-score remains constant, the model predicts that the patients’ atrophy rate at this stage would be 1.54% per year, almost three times that seen in the controls. This provides further indirect evidence to suggest that, as shown in familial AD (Schott et al, 2003) excess whole brain atrophy must be occurring prior to the onset of symptoms in sporadic disease.

Although not significant, the model predictions suggest that atrophy rate increases by 0.07% per annum per one point fall in MMSE, from a predicted rate of 1.43%yr when the MMSE is 30. In a longitudinal study of 12 patients with young onset AD (average age 47), nine of whom had familial AD, the mean rate of whole brain atrophy at the time patients had an MMSE score of 23 was calculated to be 2.8%/yr (Chan et al, 2003). Based on the results shown above, the rate of atrophy in this cohort at this
stage would be 1.95%/yr. This difference is likely to reflect the differences between familial and sporadic AD, the variability of the MMSE scores, and the inherent variability between studies. Overall, these results suggest that approximately 7-8% of inter-individual variation can be accounted for by including the baseline matrix reasoning z-score into the model; however, this had only a marginal (~10%) effect on sample size estimates.

16.5 Conclusions

The results of this work suggest that baseline level of cognitive impairment appears to predict subsequent increased rates of atrophy. This may prove to be a useful prognostic marker for AD in clinical practice, and may aid stratification in trials. A degree of caution must be exercised in drawing assumptions of causality from correlation in such analyses, and, despite the use of robust statistical methodology on large numbers of observations, it must be borne in mind that these results are based on results from a small number of subjects. Nonetheless, the results have face validity, being broadly in keeping with published reports, and suggest that further prospective studies over longer intervals are warranted to assess these factors further.
17 Regional atrophy in sporadic Alzheimer’s disease: an unbiased analysis

17.1 Introduction

To date this thesis has considered the effect of global brain atrophy in AD. As discussed previously (chapters 1 and 2), atrophy in AD not only accelerates as the disease progresses (Chan et al, 2003) but also undergoes changes in topographical distribution (Braak and Braak, 1991); thus the earliest pathological changes occur in the entorhinal cortex, progressing to involve the hippocampus, and then more lateral temporal lobe structures, cingulate and neocortex. Longitudinal imaging studies have been used to assess different brain substructures, and, using region of interest based volumetric techniques, have shown increased rates of hippocampal atrophy in established AD (Jack et al, 2000), in presymptomatic patients with familial AD (Fox et al, 1996a; Schott et al, 2003), and in subjects with mild cognitive impairment (Jack et al, 1999; Jack et al, 2004). However, such volumetric studies require a priori decisions as to which substructures should be assessed.

An alternative approach is to use automated techniques that can accurately localize regional atrophy in an unbiased manner. As discussed in chapter 2.3.1.3, one such approach involves the combination of non-linear registration of serial MRI as a means of modelling change over the whole brain in an individual, with a robust statistical assessment of areas undergoing significant change between groups. Non-linear registration is one method of determining change occurring between two scans on an individual basis. The procedure involves matching repeat scans to baseline images by using voxel-level deformation fields based on a viscous fluid model (Christensen et al, 1996; Freeborough and Fox, 1998). Jacobian determinants from this model can be used to quantify volume change at the voxel level, and colour overlays can be used to create voxel compression maps (VCMs). To determine changes that are truly disease-related, groups of individuals can be compared using statistical parametric mapping (SPM) (Friston et al, 1995), a well-validated automated technique for performing such group comparisons. Using this combination of techniques in patients with familial AD, Scahill et al demonstrated significantly increased rates of hippocampal atrophy in
presymptomatic and mildly affected patients, with a shift in the distribution of
temporal lobe atrophy with advancing disease such that the inferolateral regions of the
temporal lobes showed the most significantly increased rates of atrophy by the time
the patients were mildly or moderately affected. Significantly increased rates of
medial parietal lobe atrophy were seen at all stages, with frontal lobe involvement
occurring later in the disease (Scahill et al, 2002).

An understanding of the regional pattern and progression of atrophy in typical
sporadic AD may be useful in clinical trials: whilst rates of global atrophy may be a
useful means of determining disease modifying from symptomatic effects, an
unbiased fluid-SPM analysis, by highlighting the patterns of regional atrophy
occurring in the treated and placebo groups, may be able to determine whether such
treatments have region specific effects. This technique may also provide insights into
the disease process in AD. Whilst the results presented in chapter 16 highlight a
number of factors that predict subsequently increased rates of global brain atrophy,
using the techniques described to date, it has not been possible to determine the
regional patterns of atrophy associated with these cross-sectional predictors. The
combination of non-linear (fluid) registration and SPM provides a means of assessing
such regional changes.

The aim of this work described in this chapter was to use the combined fluid-VBM
analysis described by Scahill et al to determine significant group changes in atrophy
between patients and controls, and to extend this technique to determine the region
specific atrophy associated with the previously determined predictors of increased
atrophy.

17.2 Methods

17.2.1 MRI acquisition and pre-processing

For each individual, two serial scans acquired over the initial one-year period
(generally visits 1 to 7) were used for this analysis. In order to prevent distortion
during the non-linear (fluid) registration process, non-bias corrected images were used.
As previously, the brain was outlined using a semi-automated technique, following
which a nine degrees of freedom registration was performed to align the repeat (one-
year) scan onto the baseline image (Freeborough et al, 1996). Using this rigidly aligned repeat scan as a starting point, a non-linear (fluid) registration was carried out. The fluid registration model has been described previously by Freeborough and Fox (Freeborough and Fox, 1998) and used parameters outlined in Crum et al (Crum et al, 2001). This viscous fluid model gives an estimate of the volume change occurring at each voxel within the image, the Jacobian value. To examine the change within individuals over time, logs of the Jacobian values for each voxel within the image were analysed. In order to explore atrophy, a voxel contraction image was created (Jacobian values < 1), as described by Scahill et al (Scahill et al, 2002). Each fluid-registered image was visually assessed by two raters (Rachael Scahill and myself) to ensure adequate image quality before scans were entered into the group comparison. Where the fluid registration was marred by artefact, in some cases it was necessary to rerun the analysis from the two week scan, or to the nine month or eighteen month scans.

17.2.2 Statistical analysis

Analysis was performed using SPM99 (Wellcome Department of Cognitive Neurology, London) running in MATLAB 6 (Mathworks Incorporated, Sherbourn, Massachusetts). The images were spatially normalised to a customised template in standard stereotactic space (Talairach and Tournoux, 1998), after which they were convolved with a mask to exclude scalp, and smoothed with an isotropic Gaussian kernel of 8mm full width at half maximum. Analysis was performed using single subject, condition and covariate model with interval as a covariate, using an explicit whole brain mask to define the analysis volume. Significance levels were set at $p < 0.001$ uncorrected for multiple comparisons. To determine those areas undergoing the highest rates of change a further analysis was performed, this time corrected for multiple comparisons ($p < 0.05$).

In the first analysis, scans from all patients where available were compared to controls. In a second analysis, baseline parameters shown to be predictive of subsequently increased rates of global atrophy were assessed for their effect on regional patterns of atrophy. Patients were divided into two approximately equal groups (high and low) on the basis of the baseline characteristics, and the longitudinal fluid SPM analysis was used to assess the regional effect of both high and low scores compared to controls.
The following parameters were assessed: (1) MMSE < 20 or MMSE ≥ 20; (2) matrix reasoning z-score < -1.5 or ≥ 1.5; (3) systolic blood pressure < 140 mmHg or ≥ 140 mmHg; (4) Age at study entry < 65 years or ≥ 65 years; (5) arithmetic z-score < -2.5 or ≥ 2.5; (6) disease duration < 4 years or ≥ 4 years; (7) composite cognitive z-score < -2 or ≥ -2. Finally all analyses were repeated, covarying for matrix reasoning z-score, the only significant independent predictor of increased atrophy (chapter 16).

17.3 Results

17.3.1 Subjects

Scans from 42 patients (mean age 69.6 ± 7) and 21 controls (mean age 69.1 ± 6.8) were included in the whole group analysis. The mean inter-scan interval was 348 ± 34.9 days for the AD group and 352 ± 29.7 days for the controls. The numbers of each group included in each subgroup analysis are shown in Table 17.1

Table 17-1. Demographics of patients entering sub-group analysis.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Subgroup</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMSE</td>
<td>&lt; 20</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>≥ 20</td>
<td>18</td>
</tr>
<tr>
<td>Matrix reasoning z-score</td>
<td>&lt; -1.5</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>≥ 1.5</td>
<td>22</td>
</tr>
<tr>
<td>Arithmetic z-score</td>
<td>&lt; -2.5</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>≥ -2.5</td>
<td>20</td>
</tr>
<tr>
<td>Composite cognitive z-score</td>
<td>&lt; -2</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>≥ -2</td>
<td>22</td>
</tr>
<tr>
<td>Systolic blood pressure (mmHg)</td>
<td>&lt; 140</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>≥ 140</td>
<td>26</td>
</tr>
<tr>
<td>Age at study entry (years)</td>
<td>&lt; 65</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>≥ 65</td>
<td>23</td>
</tr>
<tr>
<td>Disease duration (years)</td>
<td>&lt; 4</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>≥ 4</td>
<td>20</td>
</tr>
</tbody>
</table>
17.3.2 Whole group analysis

Figure 17-1 shows statistical parametric maps (SPMs) in all AD subjects compared to controls. In the images uncorrected for multiple comparisons, statistically significant contraction can be seen in the AD subjects throughout the temporal lobes, with relative sparing of the superior temporal gyri and hippocampi. Atrophy is also seen in the cingulate gyrus and both parietal lobes. Atrophy is also seen in the right thalamus and ventral pons. When correction is made for multiple comparisons, the areas seen to be undergoing most significant atrophy are the medial temporal lobes, excluding the hippocampi. Some atrophy within the posterior cingulate is also seen.

17.3.3 Subgroup analyses

In the mildly affected group (MMSE ≥20), atrophy is seen almost exclusively within the medial temporal lobes, with a small amount in the cingulate (Figure 17-2). With correction for multiple comparisons, the only area showing increased atrophy compared to controls is the left hippocampus. By the time the disease is moderately advanced, there are clear increases both in the extent and focus of the increased atrophy, which now includes both temporal lobes, cingulate, precuneus, parietal lobes, with additional right thalamic, ventral pontine and frontal lobe involvement. With correction for multiple comparisons, the most significant areas undergoing atrophy are seen to be the both inferior temporal lobes, the left temporal lobe and the cingulate, with apparent prominent sparing of both hippocampi.

When the groups were separated on the basis of disease severity as assessed using the baseline matrices reasoning z-score (Figure 17-3), baseline arithmetic z-score (Figure 17-4) or baseline composite cognitive z-score (Figure 17-5), as expected, remarkably similar patterns of atrophy were seen. Thus low scores on matrix reasoning, arithmetic z-score, composite cognitive z-score (all reflecting more severe disease) all showed atrophy patterns very similar to those associated with low MMSE score. Conversely, less severe scores on all these tests resulted in patterns of atrophy very similar to those seen in patients with higher MMSE scores. Similarly, low systolic blood pressure (Figure 17-6), young age at onset (Figure 17-7), and longer disease duration (Figure 17-8) were also associated with increased global atrophy in a similar pattern to that seen with more advanced disease, i.e. a move away from medial temporal lobe
atrophy towards more lateral temporal lobe atrophy with increasing posterior cingulate involvement. Low scores on baseline MMSE, matrices reasoning z-score, composite z-score and lower baseline systolic blood pressure were all associated with significantly increased atrophy within the thalami (generally on the right) and the ventral pons; these changes were no longer statistically significant after stringent correction for multiple comparisons.

In all cases when patients were dichotomized into either high or low functioning on the basis of baseline matrix reasoning score and this was entered into the analysis as a covariate, very few areas showed significantly increased areas of atrophy compared to controls; with the more stringent analysis (i.e. correcting for multiple comparisons), no areas showed significantly increased rates of atrophy. Statistical parametric maps corrected for matrices z-score are shown for MMSE (Figure 17-9), baseline systolic blood pressure (Figure 17-10), age at disease onset (Figure 17-11) and disease duration (Figure 17-12).
Figure 17-1. Patterns of regional atrophy in sporadic Alzheimer’s disease.
Statistical parametric maps (SPMs) are shown illustrating regions undergoing significantly increased rates of atrophy in AD compared to controls (a) uncorrected for multiple comparisons; (b) corrected for multiple comparisons.
Figure 17–2. Patterns of regional atrophy by baseline MMSE.

SPMs are shown illustrating regions undergoing significantly increased rates of atrophy compared to controls (a) MMSE ≥ 20, uncorrected; (b) MMSE < 20, uncorrected; (c) MMSE ≥ 20, corrected; (d) MMSE < 20, corrected.

P<0.001, uncorrected for multiple comparisons

P<0.05, corrected for multiple comparisons
Figure 17-3. Patterns of regional atrophy by baseline matrix reasoning z-score.

SPMs are shown illustrating regions undergoing significantly increased rates of atrophy compared to controls. (a) matrices z-score ≥ -1.5, uncorrected; (b) matrices z-score < -1.5, uncorrected; (c) matrices z-score ≥ -1.5, corrected; (d) matrices z-score < -1.5, corrected.

P<0.001, uncorrected for multiple comparisons

P<0.05, corrected for multiple comparisons
Figure 17-4. Patterns of regional atrophy by baseline arithmetic z-score.

SPMs show regions undergoing significantly increased rates of atrophy compared to controls. (a) arithmetic z-score ≥ -2.5, uncorrected; (b) arithmetic z-score < -2.5, uncorrected; (c) arithmetic z-score ≥ -2.5, corrected; (d) arithmetic z-score < -2.5, corrected.

P<0.001, uncorrected for multiple comparisons

P<0.001, uncorrected for multiple comparisons

P<0.05, corrected for multiple comparisons

P<0.05, corrected for multiple comparisons
Figure 17-5. Patterns of regional atrophy by baseline composite cognitive z-score.
SPMs illustrate regions undergoing significantly increased rates of atrophy vs. controls. (a) composite cognitive z-score $\geq -2$, uncorrected; (b) composite cognitive z-score $<-2$, uncorrected; (c) composite cognitive z-score $\geq -2$, corrected; (d) composite cognitive z-score $<-2$, corrected.
Figure 17–6. Patterns of regional atrophy by baseline systolic blood pressure

SPMs illustrate regions undergoing significantly increased rates of atrophy compared to controls. (a) systolic BP ≥ 140, uncorrected; (b) systolic BP < 140, uncorrected; (c) systolic BP ≥ 140, corrected; (d) systolic BP < 140, corrected.
Figure 17–7. Patterns of regional atrophy by age at disease onset.

SPMs illustrate regions undergoing significantly increased rates of atrophy compared to controls. (a) age at onset ≥ 65, uncorrected; (b) age at onset < 65, uncorrected; (c) age at onset ≥ 65, corrected; (d) age at onset < 65, corrected.
Figure 17–8. Patterns of regional atrophy by disease duration. SPMs are shown illustrating regions undergoing significantly increased rates of atrophy compared to controls. (a) disease duration < 4 years, uncorrected; (b) disease duration ≥ 4 years, uncorrected; (c) disease duration < 4 years, corrected; (d) disease duration ≥ 4 years, corrected.
Figure 17-9. Regional atrophy by MMSE corrected for matrices z-score.

SPMs are shown illustrating regions undergoing significantly increased rates of atrophy compared to controls after correction for baseline matrices z-score. (a) Baseline MMSE ≥20, uncorrected; (b) MMSE < 20, uncorrected; (c) MMSE ≥ 20, corrected; (d) MMSE < 20, corrected.
Figure 17–10. Regional atrophy by systolic BP corrected for matrices z-score.

SPMs illustrate regions undergoing significantly increased rates of atrophy compared to controls after correction for baseline matrices z-score. (a) systolic BP ≥ 140, uncorrected; (b) systolic BP < 140, uncorrected; (c) systolic BP ≥ 140, corrected; (d) systolic BP < 140, corrected.
Figure 17–11. Regional atrophy by age at AD onset corrected for matrices z-score.

SPMs illustrate regions undergoing significantly increased rates of atrophy compared to controls after correction for baseline matrices z-score. (a) age at onset ≥ 65, uncorrected; (b) age at onset < 65, uncorrected; (c) age at onset ≥ 65, corrected; (d) age at onset < 65, corrected.
Figure 17–12. Regional atrophy by AD duration corrected for matrices z-score.

SPMs illustrate regions undergoing significantly increased rates of atrophy after correction for baseline matrices z-score. (a) disease duration < 4 years, uncorrected; (b) disease duration ≥ 4 years, uncorrected; (c) disease duration < 4 years, corrected; (d) disease duration ≥ 4 years, corrected.
Discussion

These results demonstrate that this technique that combines assessment of regional atrophy on an individual basis with a robust statistical group analysis, provides a means of assessing areas undergoing significantly increased rates of atrophy. When all patients were compared to controls, the observed areas undergoing significantly increased regional atrophy were consistent with the known histopathological (Braak and Braak, 1991) and imaging (de Leon et al, 1993; Scahill et al, 2002; Schott et al, 2003) staging of the disease; thus the disease initially involves the medial temporal lobes and posterior cingulate, before spreading more laterally within the temporal lobes before involving the parietal and frontal lobes.

The subgroup analysis showed increasing atrophy with increasing severity, as predicted by the multi-level model; consistent patterns of evolving regional atrophy were seen using a number of measures of disease severity (i.e. MMSE, matrix reasoning, arithmetic, and a composite cognitive score). It is notable that in the MMSE analysis, very little atrophy survives correction for the alternative measure of severity, matrix reasoning z-score. This suggests that as predicted by the multi-level model, the MMSE is providing little, if any, additional information beyond that associated with the matrix reasoning z-score. The observed atrophy patterns with increasing disease severity are comparable to those from a study of familial AD using this approach (Scahill et al, 2002); thus increasing severity of disease is associated with both more brain areas undergoing significant atrophy, and a move from medial to more lateral temporal atrophy, with extension to neocortical areas as the disease moves from mild to moderate stages (Smith, 2002).

As predicted by the multi-level model, younger age at onset was associated with an increased rate of atrophy in a similar pattern to that seen with more severe disease, providing further evidence that the effect of age on atrophy is likely to be mediated via disease severity (see chapter 16.4). Further evidence for this association is provided by the results seen when this analysis is performed corrected for baseline matrix reasoning z-score, when virtually no significant excess atrophy is observed (Figure 17-11). As discussed in chapter 16.4, these findings might suggest a different phenotypic presentation of AD at a younger age; alternatively it is perhaps more likely that younger patients with
more advanced AD and their carers are more likely to participate in studies than older patients with similar disease severity.

The conclusion that lower systolic blood pressure is a consequence of increasing disease severity (see chapter 16), is also supported by the imaging findings: the pattern and progression of atrophy associated with decreasing blood pressure was similar to that associated with increasing disease severity, and virtually no excess atrophy was seen when correction was made for severity. This provides yet further evidence to suggest that lower systolic blood pressure is a consequence and not a cause of more advanced AD (Skoog et al, 1998).

An interesting and unexpected finding was the observation of atrophy within the ventral pons in both the more severely affected patients and those with lower blood pressure. Although atrophy was not seen with the stringent correction for multiple comparisons, assessing regional changes in the brainstem presents particular difficulties using this technique, and artefactual problems cannot be excluded, it is interesting that histological studies have demonstrated that the pontine parabrachial nuclei are prominently affected in AD, and increasingly so during the progression of the disease (Rub et al, 2001). These nuclei are crucially involved in autonomic circuits modifying cardiovascular function (Rub et al, 2001). Although the focus of atrophy suggested by this study is ventral to these nuclei, it is tempting to speculate that the demonstrated atrophy pattern may be linked to the histopathological involvement of these nuclei; this in turn could provide an explanation not only for the finding that in this study systolic hypotension was a function of disease severity, but also for the consistent finding of relative hypotension associated with the progression of AD (Davis et al, 2003; Skoog et al, 1996).

As well as providing insights into the progression of AD, this technique may also be useful in clinical trials. As previously discussed, if a trial drug were to reduce rates of global atrophy compared to placebo, this would provide evidence to suggest a disease modifying effect. Scans acquired during such a trial could then be analyzed using the combined fluid-SPM technique; comparing the treated and untreated cohorts and determining differences between them could provide an unbiased means of determining a region specific drug effect. This could provide further qualitative evidence for drug efficacy, as well as providing insights into the mechanism of drug action.
18 Conclusions

The advent of symptomatic therapies for the treatment of AD has been a major advance in the management of this devastating and increasingly common condition. The prospect of disease-slowing therapies has created an urgent need for reliable surrogate markers of disease progression. Global brain atrophy has considerable potential in this area, and short interval scanning may be a feasible means of determining disease modifying from symptomatic effects, whilst at the same time reducing the number of patients required in a trial and the length of time over which experimental drugs must be studied. This in turn may reduce both the cost of drug development, and the length of time taken for drugs to be made available to patients. The purpose of this thesis was to assess the feasibility of using short interval scanning as a surrogate marker of disease progression in AD by addressing a number of specific aims outlined in chapter 3.3.5.

18.1 Can patients tolerate serial scans?

The results presented in this thesis suggest that patients with mild-moderate disease are able to tolerate multiple short-interval serial MRI scans. Despite the very involved nature of this study (7 scanning visits in 12 months), only 1 patient dropped-out, ~97% of planned visits were achieved, ~94% of planned scans were acquired and usable, and ~96% of these scans were of good quality. Whilst drop-out rates are likely to be greater in multi-centre trials, particularly if trial drugs are associated with side effects, these data suggest that serial scanning is feasible, at least in a specialist setting. Strategies that may have contributed to this high retention rate and good scan quality include: (1) establishing a good rapport with patient and carer, providing realistic information about the purpose and demands of the trial (including specific dates and times for all scanning visits one year in advance) and providing a single trial coordinator as a point of contact; (2) providing a safe and efficient means of transport to and from the scans and ensuring that patients were met by a familiar face at each scanning visit; (3) performing all scans on the same scanner, with the same radiographer alert to the specific needs of patients with AD; and (4) ensuring that the scan protocol was as short as possible to maximize compliance. These simple factors which could be emphasized to and emulated by trial coordinators in treatment trials may serve to improve the chances of retaining patients in studies and acquiring high quality images.
18.2 Is 9dof registration a valid means of correcting for voxel drift?

It is widely accepted that registration, by accurately aligning scans, allows for more accurate comparison of serially acquired images. An inherent problem with acquiring serial scans is the unpredictable and inevitable scanner changes that occur between the two scans. Such changes, as demonstrated in this study (chapter 9) can result in apparent atrophy changes of ~2%, i.e. the yearly magnitude of change seen in AD. 9dof registration allows for a degree of voxel stretch to overcome this problem, but comes with a theoretical risk that atrophy can be "corrected away". The results presented in chapter 9 demonstrate conclusively that 9dof registration is able to correcting for substantial changes in voxel dimension drift, and to do so without altering measures of true atrophy in AD. Furthermore, the 9dof registration was able to correct these changes at least as accurately as the very time consuming manual alternative of correcting for total intracranial volume. It is therefore concluded that 9dof registration is a valid, automated method allowing for accurate comparison of serial scans.

18.3 Does altering the BBSI parameters alter calculated atrophy rates?

The BBSI is a semi-automated validated method for determining whole brain atrophy from registered scan pairs. Two parameters, the window width and window centre must be set by the operator; previous studies have used single scans or scan normalization processes to determine these parameters. In this study, by critically assessing the effect of altering these parameters, it was possible to demonstrate an improvement in the accuracy of whole brain atrophy measurement without reducing precision. The methodology underlying the choice of parameters can be implemented in any serial data-set, and may reduce the sample sizes required in clinical trials.

18.4 Can atrophy rates distinguish patients from controls at short intervals?

At the time of writing, only one published study has reported atrophy rates in AD at intervals of less than one year (Bradley et al, 2002). The results presented in chapters 10 and 12 of this thesis demonstrate that significant differences between patients with AD and age-matched controls can be reliably determined at intervals as short as six months. Furthermore the results of the BBSI analysis are unaffected whether or not only "good" or
all available scans are used for analysis, demonstrating the robustness of the technique, and suggesting that it may be useful in larger multi-centre studies where less reliable scan quality might be expected. The finding that excess atrophy in AD can be determined at short intervals provides a signal that could be harnessed in clinical trials or for diagnostic purposes.

18.5 Are ventricular or BBSI derived atrophy rates better surrogate markers?

The results of a direct comparison between ventricular and BBSI derived atrophy rates (chapter 12.2.1) suggested that in a simple two time point six month trial, ~2.7 (401 vs 170) times as many patients would be required in each arm of a placebo-controlled study using the BBSI compared to ventricular measures of atrophy. This difference is considerably smaller (139 vs 150 i.e. ~11%) at one year. The key determinant of these differences is the variance of the measurement. The ventricles are centrally positioned large structures with well defined borders, and are less susceptible to scanner inhomogeneity, and are therefore less variable than measures of whole brain, especially over short intervals. These results would suggest that, particularly at short intervals, ventricular measures would be better surrogate markers of AD progression than the direct BBSI measure. There are however a number of caveats with the use of ventricular measures of atrophy: (1) ventricular measures require manual tracing by trained raters with associated time and cost implication; (2) many other factors can result in significant changes in ventricular size without a corresponding change in brain volume; and (3) it has been suggested that ventricular change in AD is correlated but not entirely accounted for by AD neuropathology (Silbert et al, 2003). Thus whilst ventricular enlargement may be a more powerful means of determining change, it may also be a less biological plausible surrogate than whole brain atrophy. One potential approach might be to combine the results of these two techniques together in some way, thus combining the biological plausibility of direct measures of brain atrophy with the increased statistical power that may be gained from ventricular measures.

18.6 Can BBSI derived atrophy rates be mathematically modelled?

The key determinant of statistical power is the variance of the measurement. An accurate understanding of the variance and source of variance in atrophy rates then allows for a critical evaluation of different trial designs. As demonstrated in Table 13-1, atrophy
measured over several different six month intervals, even in essentially the same cohort, can result in large variations in mean atrophy rate (1.93% to 2.47% per year), and still larger variations in the measured standard deviation of atrophy rate (1.5% to 2.4%). In this study, by virtue of the multiple scans acquired for each individual, multiple direct measures of atrophy could be calculated over a wide range of intervals. Thus for the AD cohort, 2199 direct, automated (BBSI) measures were made, and 1182 for the controls. The multi-level model specifically designed to analyse this study was implemented, and estimates for mean, variance and sources of variance were then determined. For the patients, the mean atrophy rate was determined to be 2.23% per year, in line with previous reports. The standard deviation at one year was determined to be 1.28%, with inter-individual variance accounting for approximately ¾ of the total. For the controls, at one year the mean rate of atrophy was 0.49% per year, with a standard deviation of 0.6% which appeared to be entirely accounted for by intra-individual variability.

The close comparison between the estimated and observed atrophy rate demonstrate that a multi-level mathematical model is able to model the atrophy in AD with a high degree of accuracy. By defining not only the mean and variance of atrophy, but also the sources of this variance, the model provides a powerful means of investigating different trial designs; furthermore it provides a means of assessing some of the potential determinants of inter- and intra-individual variation. This modelling approach can easily be applied to other serially acquired direct measures of atrophy, and may therefore be useful in the analysis of future experimental and trial data.

18.7 Are there advantages in combining scan intervals for trials?

Using the model-determined estimates of mean, variance and sources of variance in atrophy rates it is possible to estimate the sample size requirements for trials of varying duration involving two or more scan visits. The results determined in chapter 14 suggest a number of factors that influence such studies: these include: (1) length of proposed trial; (2) numbers of scans included in the trial design; and (3) the timing of these scans. For trials over long periods (e.g. two years), little practical benefit is seen by adding additional scans; however, as the trial interval decreases, substantial reductions in sample sizes may be achieved by increasing the number of scans acquired. Thus over six months, an estimated 43% reduction in sample sizes may be achieved by acquiring seven serial scans.
instead of two. The timing of these scans is also critical: maximum benefit is seen when additional scans are placed close to either the start or end of the study.

Whilst these estimates provide useful information regarding trial designs, in the context of a pharmaceutical trial a number of other factors must be taken into account. The expected effect of the drug will critically affect any sample sizes estimates – in this thesis a 20% drug effect was modelled; a larger effect dramatically reduces sample sizes, and a smaller effect requires an increase. The effect of scan drop-outs not only increases sample sizes, but also may be related to the length of any trial design. Acquiring multiple scans over short intervals, by increasing contact with the trial centre, might actually increase retention in the trial. Furthermore, multiple scan trial designs may also allow for useful information to be gained from patients dropping out before the end of the study, but after two or more scans have been acquired. There is therefore a balance to be struck between placing scans at the beginning and end of a study to increase power, and placing additional scans in the middle of the study which may provide important information in the event of patients dropping out prior to completion of the study. Finally any trial design must take into account the expected relationship between dosing and drug effect; clearly if a drug has a substantial lag period, shorter trial designs may not capture the full treatment effect.

18.8 What are the correlates and predictors of increased atrophy?

Whilst the relatively small number of patients in this study precluded a standard assessment of factors influencing and correlating with change, the large numbers of observations per patient and the robust statistical framework provided by the multi-level model do allow some exploratory investigation. The results presented in chapter 16 suggest that change in a number of non-memory based measures of cognition (naming, visuoperceptual skills and a composite non-memory cognitive score) correlated with increased rates of cerebral atrophy, and thus provide further evidence to support whole brain atrophy as a valid, biological plausible surrogate marker of AD progression. The results from chapter 16 suggest that, of the variables tested, severity of baseline cognitive impairment was the single best predictor of increased rates of atrophy. Age at onset and low systolic blood pressure were also predictive of increased rates of atrophy, although there was evidence to suggest that these were both manifestations of disease severity. Whilst these effects are all biological plausible and provide some insights into factors
influencing the progression of AD, it is also interesting to note that the single best (and only independent) predictor of atrophy (the matrix-reasoning score) only accounted for 7-8% of inter-individual variability in atrophy rate. It is thus clear that other factors, of which one or many may not yet be known, must account for the majority of such variability; this methodology provides one means in which potential influences can be tested. The determination of other predictors of increased atrophy may in turn increase our understanding of the pathogenesis and progression of AD, and, if some of these factors are modifiable, may have direct relevance to patient care.

18.9 What patterns of regional atrophy are seen in sporadic AD?

The combination of an unbiased methodology for detecting change at the individual level (fluid registration) with a robust means of comparing groups (SPM) provides a powerful means of assessing patterns of regional atrophy. The results from this study demonstrate that, as predicted by the multi-level model, disease severity has a major influence on both the extent and topography of atrophy in AD; furthermore the observed patterns were in keeping with the published data on the histopathological and imaging progression of AD. This methodology may not only be useful in delineating the patterns of atrophy in AD, but may prove to be a useful adjunctive measure in clinical trials in AD, providing complementary, qualitative evidence of any region specific drug effects.

The combination of fluid registration and SPM provided further evidence to support the findings that the increased atrophy associated with younger age at onset and lower systolic blood pressure were mediated via disease severity. The intriguing, although speculative, finding of ventral pontine atrophy associated with lower systolic blood pressure and more advanced disease may suggest an anatomical basis for a link between hypotension and AD. Similarly the association between more advanced AD and lower systolic blood pressure raised questions with regard to the practical management of patients with AD, who commonly are prescribed antihypertensive drugs. Larger prospective studies are required to address these questions fully.
18.10 Possible future applications

The methodologies described in this thesis may find useful application in further studies in neurodegeneration or other brain disorders. Atrophy derived from serial MR scans may be a useful means of assessing novel genetic or environmental factors that might influence the development or progression of Alzheimer's disease, and for tracking the progression of other neurodegenerative diseases. Should specific disease modifying drugs for different dementias become available, early and accurate diagnosis will become increasingly important; serial imaging demonstrating excess atrophy above that expected for age may prove to be a useful means of detecting early neurodegeneration and the pattern of regional atrophy may prove to be a useful means of distinguishing different dementias from one another.

18.11 Summary

Serial MR imaging allows for rates of brain atrophy to be calculated in vivo in Alzheimer’s disease. Atrophy rates are feasible, biologically plausible markers of progression which may be useful in clinical trials in AD. Atrophy in AD may be modelled mathematically, and the results obtained from this analysis may be used to estimate sample sizes required in trials of varying duration, and varying numbers of scans. The results from this study suggest that, using whole brain atrophy derived from serial MRI scans as an outcome measure, drug trials to assess progression may be feasible over relatively short intervals (e.g. six months), and that there may be several advantages in acquiring multiple scans during such trials. The combination of fluid registration and SPM may be a useful methodology to demonstrate region specific drug effects, and thus be a useful adjunct to measures of atrophy in such trials. However, definitive validation of atrophy as a true surrogate marker of AD must await the development of a truly disease-modifying agent.
Publications

Publications arising from this thesis


Schott JM, Simon JE, Whitwell JL, MacManus DG, Frost C, Wang, L, Bartlett PA, Boyes RG, Rossor MN, Fox NC. Global brain atrophy as a surrogate marker of progression in Alzheimer’s disease: a one-year, prospective, longitudinal MRI study using the brain boundary shift integral. Neurology 2003; 60(S1); A161 (abstract)

Papers related to this thesis


Acknowledgements

I am indebted to all the patients, control subjects and carers who took part in this study. Their dedication, often under extremely difficult circumstances, made this work possible.

This thesis would not have been possible without the support, expertise and help of numerous people within the Dementia Research Group and Institute of Neurology. Drs Jessica Simon and Liqun Wang were instrumental in piloting this study. Sebastian Crutch performed all the neuropsychological assessments, and Professor Elizabeth Warrington was integral in designing the neuropsychological battery and assessing ANC (chapter 7). The majority of the scan segmentations were performed by Rhian Jenkins and Jenny Whitwell, who also measured the TIVs for chapter 9. Dr Emma Lewis provided physics advice and performed many of the experiments described in chapter 9. I would like to thank Shona Price for both performing and teaching me ventricular segmentation (chapter 10). Richard Boyes helped design and create the scan database; he also performed much of the BBSI validation work described in chapter 11. Chris Frost provided statistical advice throughout; I am particularly grateful for his help with the multi-level model (chapters 13-16), which he designed and implemented. Dr Rachael Scahill provided expertise in all aspects of fluid registration and SPM (chapter 17). Mr David MacManus performed all the MRI scans, Dr John Stevens provided neuroradiological advice, Dr Dan Healy and John Beck performed the ApoE analyses, and Dr Angus Kennedy (Chelsea and Westminster hospital) referred me a number of patients. Professor Martin Rossor diagnosed most of the patients described in this thesis, and is a role model for combining skilled patient care with first class research.

I am indebted to my supervisor, Dr Nick Fox, who as well as conceiving this project, provided outstanding support, advice, enthusiasm and friendship throughout its execution.

Funding for this work was provided by the Alzheimer’s Society, and by GlaxoSmithKline.

I would like to thank Anette Schrag, Benjamin, Judith and Geoffrey Schott who put up with me during the writing of this thesis, provided endless moral and practical support, and continue to make it all worthwhile.
Appendix 1. Criteria for the diagnosis of dementia

Diagnostic and Statistical Manual of Mental Disorders (4th Edition) (DSM-IV)

I. The development of multiple cognitive deficits manifested by both:

a) Memory impairment
   *impaired ability to learn new information or recall learned information*

b) and one (or more) of the following cognitive disturbances:

(i) Aphasia
   *language disturbance*

(ii) Apraxia
    *impaired ability to carry out motor activities despite intact motor function*

(iii) Agnosia
    *failure to recognize or identify objects despite intact sensory function*

(iv) Disturbance in executive functioning
    *i.e., planning, organizing, sequencing, abstracting*

II. The cognitive deficits in Criteria Ia and Ib each cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning

III. The deficits do not occur exclusively during the course of a delirium

Dementia is diagnosed if criteria I, II and III are met

(Note that only one of the features listed under Ib has to be present)
Appendix 2. Criteria for the diagnosis of Alzheimer’s disease

National Institute of Neurological and Communicative Disorders and the Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) (McKhann et al, 1984)

Probable Alzheimer’s disease (if all criteria I are answered with “yes”)

I. The diagnosis of PROBABLE Alzheimer’s disease is supported by:
   a) Dementia established by clinical examination and documented by the MMSE; Blessed Dementia Scale, or similar examination, and confirmed by neuropsychological tests
   b) Deficits in two or more areas of cognition
   c) Progressive worsening of memory and other cognitive functions
   d) No disturbance of consciousness
   e) Onset between ages 40 and 90, most often after age 65
   f) Absence of systemic disorders or other brain diseases that in and of themselves could account for the progressive deficits in memory and cognition

II. The diagnosis of PROBABLE Alzheimer’s disease is supported by:
   a) Progressive deterioration of specific cognitive functions such as language (aphasia), motor skills (apraxia), and perceptions (agnosia)
   b) Impaired activities of daily living and altered patterns of behaviour
   c) Family history of similar disorders, particularly if confirmed neuropathologically
   d) Normal lumbar puncture as evaluated by standard techniques
   e) Normal pattern or non-specific EEG changes, e.g. increased slow-wave activity
   f) Evidence of cerebral atrophy on CT with progression documented by serial observation

III. Other clinical features consistent with PROBABLE Alzheimer’s disease:
   a) Plateaus in the course of progression of the illness
   b) Associated symptoms of depression, insomnia, incontinence, delusions, illusions, hallucinations, catastrophic verbal, emotional, or physical outbursts, sexual disorders, and weight loss
c) Other neurological abnormalities in some patients, especially with more advanced
disease and including motor signs such as increased muscle tone, myoclonus, or gait
disorder
d) Seizures in advanced disease
e) CT normal for age

IV. Features that make PROBABLE Alzheimer's disease uncertain or unlikely:
a) Sudden, apoplectic onset
b) Focal neurological findings such as hemiparesis, sensory loss, visual field deficits,
and incoordination early in the course of the illness
c) Seizures or gait disturbances at the onset or very early in the course of the illness

Possible Alzheimer's disease

I. May be made on the basis of the dementia syndrome, in the absence of other
neurological, psychiatric, or systemic disorders sufficient to cause dementia, in the
presence of variations in the onset, in the presentation, or in the clinical course.

II. May be made in the presence of a second systemic or brain disorder sufficient to
produce dementia, which is not considered to be the cause of the dementia.

III. Should be used in research studies when a single, gradually progressive severe
cognitive deficit is identified in the absence of other identifiable cause

Definite Alzheimer's disease

Definite AD refers to histopathologically confirmed disease. The methods by which
pathologists should make this diagnosis are described in Mirra et al (Mirra et al, 1991).
Histopathological methods (and AD criteria) are imperfect and continue to be refined but
in essence depend on demonstrating sufficient age-related densities of neurofibrillary
tangles and amyloid plaques (Newell et al, 1999).
Appendix 3. Mini-Mental State Examination (MMSE)

“What is the: Year? Season? Date? Day of the week? Month?”
Score one for each correct answer (max 5)

“What is the name of this: Country? City? Area? Hospital? What floor are we on?”
Score one for each correct answer (max 5)

“Please repeat these three objects after me, and remember them: Apple, Penny, Table”
Score one for each item correctly repeated on the first attempt (max 3)

“Please count backwards from 100 in steps of 7” (Stop after five answers) or
“Please spell the word WORLD backwards”
Score one for each correct response (max 5)

“Please recall the three objects I asked you to remember”
Score one for each item (Apple, Penny, Table) correctly recalled (max 3)

“Please name the following items” (show a watch and a pen)
Score one for each item correctly identified (max 2)

“Please repeat after me “no ifs, ands, or buts” ”
Score one if sentence repeated correctly on the first attempt (max 1)

“Please take this paper in your right hand, fold it in half, and put it on the floor”
Score one for each of the three stages completed correctly (max 3)

“Please follow the instruction: Show card reading "CLOSE YOUR EYES" ”
Score one if patient closes eyes (max 1)

“Please write a sentence”
Score one point if sentence has a subject, a verb, and makes sense (max 1)

“Please copy this picture (show drawing of a pair of intersecting pentagons)”
Score one point if performed correctly (max 1)

Maximum score 30
## Appendix 4. Clinical Dementia Rating scoring sheet

<table>
<thead>
<tr>
<th>Impairment Level and CDR Score (0, 0.5, 1, 2, 3)</th>
<th>None: 0</th>
<th>Questionable: 0.5</th>
<th>Mild: 1</th>
<th>Moderate: 2</th>
<th>Severe: 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Memory</strong></td>
<td>No memory loss or slight inconsistent forgetfulness</td>
<td>Consistent slight forgetfulness; partial recollection of events; &quot;benign&quot; forgetfulness</td>
<td>Moderate memory loss; more marked for recent events; defect interferes with everyday activities</td>
<td>Severe memory loss; only highly learned material retained; new material rapidly lost</td>
<td>Severe memory loss; only fragments remain</td>
</tr>
<tr>
<td><strong>Orientation</strong></td>
<td>Fully oriented</td>
<td>Fully oriented except for slight difficulty with time relationships</td>
<td>Moderate difficulty with time relationships; oriented for place at examination; may have geographic disorientation elsewhere</td>
<td>Severe difficulty with time relationships; usually disoriented to time, often to place</td>
<td>Oriented to person only</td>
</tr>
<tr>
<td><strong>Judgment &amp; Problem Solving</strong></td>
<td>Solves everyday problems &amp; handles business &amp; financial affairs well; judgment good in relation to past performance</td>
<td>Slight impairment in solving problems, similarities, and differences</td>
<td>Moderate difficulty in handling problems, similarities, and differences; social judgment usually maintained</td>
<td>Severely impaired in handling problems, similarities, and differences; social judgment usually impaired</td>
<td>Unable to make judgments or solve problems</td>
</tr>
<tr>
<td><strong>Community Affairs</strong></td>
<td>Independent function at usual level in job, shopping, volunteer and social groups</td>
<td>Slight impairment in these activities</td>
<td>Unable to function independently at these activities although may still be engaged in some; appears normal to casual inspection</td>
<td>No pretense of independent function outside home. Appears well enough to taken to functions outside a family home</td>
<td>No pretense of independent function outside home. Appears too ill to be taken to functions outside family home</td>
</tr>
<tr>
<td><strong>Home and Hobbies</strong></td>
<td>Life at home, hobbies, and intellectual interests well maintained</td>
<td>Life at home, hobbies, and intellectual interests slightly impaired</td>
<td>Mild but definite impairment of function at home; more difficult chores abandoned; more complicated hobbies and interests abandoned</td>
<td>Only simple chores preserved; very restricted interests, poorly maintained</td>
<td>No significant function in home</td>
</tr>
<tr>
<td><strong>Personal Care</strong></td>
<td>Fully capable of self-care</td>
<td>Needs prompting</td>
<td>Requires assistance in dressing, hygiene, keeping of personal effects</td>
<td>Requires much help with personal care; frequent incontinence</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 5. Modified Hachinski Ischaemic Score

(Rosen et al, 1980)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Code</th>
<th>x</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrupt onset</td>
<td>0</td>
<td>x2</td>
<td></td>
</tr>
<tr>
<td>Stepwise deterioration</td>
<td>0</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>Somatic complaints</td>
<td>0</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>Headache, tinnitus, chest pain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional incontinence</td>
<td>0</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>Laughing or crying with minimal provocation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of hypertension</td>
<td>0</td>
<td>x1</td>
<td></td>
</tr>
<tr>
<td>BP&gt;160/100 on 3 occasions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Or requiring diet modification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of strokes</td>
<td>0</td>
<td>x2</td>
<td></td>
</tr>
<tr>
<td>Aphasia, hemiparesis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focal neurological symptoms</td>
<td>0</td>
<td>x2</td>
<td></td>
</tr>
<tr>
<td>Transient monocular blindness, seizures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unilateral weakness/sensory loss</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focal neurological signs</td>
<td>0</td>
<td>x2</td>
<td></td>
</tr>
<tr>
<td>Asymmetric rigidity or reflexes,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extensor plantars, nystagmus</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total ischaemic score

221
Appendix 6. Whole brain segmentation protocol

Whole brain segmentation was performed using a protocol described in detail by Freeborough et al (Freeborough et al, 1997)

1. An upper threshold value is set to ensure no brain tissues are outlined (Figure A 1).

Figure A 1. Setting the upper threshold

2. The lower threshold is then set to a value to include brain but not CSF. The aim is to outline the brain as accurately as possible (Figure A 2).

Figure A 2. Setting the lower threshold
3. The axial cut off is set to remove any non-brain below the cerebellum; this provides a reproducible lower/inferior limit (Figure A 3)

Figure A 3. Defining the inferior limit of the scan.

4. A series of erosions are applied to “eat away” at the external surfaces of the brain region that has been outlined above until only brain remains (see Figure A 4).

Figure A 4. Using erosions to define brain tissue.
5. A series of dilations are applied to recover the brain that has been removed by the erosion process (see Figure A 5) while excluding non-brain voxels.

Figure A 5. Using dilations to recover lost brain tissue.

6. The final step is to apply a rethresholding box to “fill in” any area within that region that has been omitted by step 5. Subsequently manual editing may be required to ensure that the brain is correctly outlined.
Appendix 7. Ventricular segmentation protocol

The aim of the protocol is to label areas of ventricular cerebrospinal fluid, but not extraneous CSF spaces. Using the segmented whole brain (see Appendix 6), a threshold of 60% mean voxel intensity value from the whole brain is used to define the ventricular spaces. "Seeds" are implanted within the ventricles, and the outlined regions are propagated caudally and rostrally to include the lateral ventricles. Manual editing is required to ensure that extraneous CSF spaces are not included, that the temporal horns are included, and that the third but not fourth ventricle is outlined (see Figure A 6).

Figure A 6. Ventricular segmentation.
Sagittal (upper) and coronal (lower) views.
<table>
<thead>
<tr>
<th>Years unwell</th>
<th>Age at symptoms</th>
<th>CDR-SB</th>
<th>CDR</th>
<th>ApoE dose</th>
<th>Hachinski score</th>
<th>BMI (kg/m²)</th>
<th>Diastolic BP (mmHg)</th>
<th>Systolic BP (mmHg)</th>
<th>School leaving age</th>
<th>Start MMSE</th>
<th>Age at start</th>
<th>Male</th>
<th>Patient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>52</td>
<td>63</td>
<td>17</td>
<td>29</td>
<td>88.8</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>30</td>
<td>85</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>0</td>
<td>22</td>
<td>70</td>
<td>17</td>
<td>49</td>
<td>63.4</td>
<td>72.9</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>18</td>
<td>19</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Patient characteristics
<table>
<thead>
<tr>
<th>Id</th>
<th>Patient</th>
<th>Male</th>
<th>Age at start</th>
<th>Start MMSE</th>
<th>School leaving age</th>
<th>Systolic BP (mmHg)</th>
<th>Diastolic BP (mmHg)</th>
<th>BMI (kg/m²)</th>
<th>Hachinski score</th>
<th>ApoE dose</th>
<th>CDR</th>
<th>CDR-SB</th>
<th>Age at symptoms</th>
<th>Years unwell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years unwell</td>
<td>Age at symptoms</td>
<td>CDR-SB</td>
<td>CDR</td>
<td>ApoE dose</td>
<td>Hachinski score</td>
<td>BMI (kg/m²)</td>
<td>Diastolic BP (mmHg)</td>
<td>Systolic BP (mmHg)</td>
<td>Diagnosed MMSE</td>
<td>Start MMSE</td>
<td>Age at start</td>
<td>Male</td>
<td>Patient Id</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------</td>
<td>--------</td>
<td>-----</td>
<td>------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>----------------</td>
<td>------------</td>
<td>-------------</td>
<td>------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>96</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>25</td>
<td>144</td>
<td>0.0</td>
<td>17</td>
<td>26</td>
<td>×</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>3.9</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>22</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>×</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>4.1</td>
<td>74</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>66</td>
<td>17</td>
<td>18</td>
<td>21</td>
<td>15</td>
<td>×</td>
<td>69</td>
<td></td>
</tr>
<tr>
<td>4.8</td>
<td>60</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>6.8</td>
<td>29</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>60</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>89</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td>3.9</td>
<td>50</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>89</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>89</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>70</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>25</td>
<td>14</td>
<td>18</td>
<td>18</td>
<td>18</td>
<td>×</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>4.8</td>
<td>78</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>×</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>65</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>72</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>75</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>5.2</td>
<td>72</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>99</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>95</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>72</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>4.7</td>
<td>63</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>22</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>5.1</td>
<td>72</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>23</td>
<td>06</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>×</td>
<td>95</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 9. Pre-scan questionnaire

We would be grateful if you, or someone who knows you well, could fill in this form before your next scan. Please bring the form with you when you come to visit us. If you are unsure how to answer any of the questions, we can help you when you are here.

Name........................................................................................................ Date ...................

1. Has there been any change in the medicines that you take, since your last visit?
   Yes □
   No □
   If yes, please detail with the medicine name, the dose and the rough date of change:

2. Have you had any new health problems, illnesses or operations since your last visit?
   Yes □
   No □
   If Yes, please detail, however minor:

3. Has there been a significant change in your memory or thinking since your last scan?
   Yes □
   No □
   If Yes, please detail:

4. Have you had a head injury since your last scan? Yes □ No □

5. Have you had a fit or seizure since your last scan? Yes □ No □

6. Have you had a pacemaker inserted, or had any other implant or an injury such that metal may have entered your body? Yes □ No □
References


56. Fox NC, Seahill RI, Crum WR, Rossor MN. Correlation between rates of brain atrophy and cognitive decline in AD. Neurology 1999a; 52: 1687-1689.


239


