Assessment and optimisation of MRI measures of atrophy as potential markers of disease progression in multiple sclerosis

Thesis submitted for the degree of

Doctor of Philosophy

Valerie Anderson

UCL

Institute of Neurology

2008

Declaration statement

I, Valerie Anderson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis and acknowledgements.

Valerie Anderson

Abstract

There is a need for sensitive measures of disease progression in multiple sclerosis (MS) to monitor treatment effects and understand disease evolution. MRI measures of brain atrophy have been proposed for this purpose. This thesis investigates a number of measurement techniques to assess their relative ability to monitor disease progression in clinically isolated syndromes (CIS) and early relapsing remitting MS (RRMS).

Presented, is work demonstrating that measurement techniques and MR acquisitions can be optimised to give small but significant improvements in measurement sensitivity and precision, which provided greater statistical power. Direct comparison of numerous techniques demonstrated significant differences between them. Atrophy measurements from SIENA and the BBSI (registration-based techniques) were significantly more precise than segmentation and subtraction of brain volumes, although larger percentage losses were observed in grey matter fraction. Ventricular enlargement (VE) gave similar statistical power and these techniques were robust and reliable; scan-rescan measurement error was <0.01% of brain volume for BBSI and SIENA and <0.04ml for VE.

Annual atrophy rates (using SIENA) were -0.78% in RRMS and -0.52% in CIS patients who progressed to MS, which were significantly greater than the rate observed in controls (-0.07%). Sample size calculations for future trials of disease-modifying treatments in RRMS, using brain atrophy as an outcome measure, are described. For SIENA, the BBSI and VE respectively, an estimated 123, 157 and 140 patients per treatment arm respectively would be required to show a 30% slowing of atrophy rate over two years. In CIS subjects brain atrophy rate was a significant prognostic factor, independent of T2 MRI lesions at baseline, for development of MS by five year follow-up. It was also the most significant MR predictor of disability in RRMS subjects. Cognitive assessment of RRMS patients at five year follow-up is described, and brain atrophy rate was a significant predictor of overall cognitive performance, and more specifically, of performance in tests of memory.

The work in this thesis has identified methods for sensitively measuring progressive brain atrophy in MS. It has shown that brain atrophy changes in early MS are related to early clinical evolution, providing complementary information to clinical assessment that could be utilised to monitor disease progression.

Contents

Figures

Tables

Problem and aims

The identification and provision of effective new disease-modifying therapies for people with multiple sclerosis (MS) are key research and public health aims. These aims have in turn increased the need to improve the assessment of disease progression and the measurement of therapeutic effects. Clinically-stable individuals with MS often have underlying disease activity, including progressive loss of myelin and nerve fibres in the central nervous system, which means that clinical measures are an insensitive way of monitoring disease activity and detecting disease-slowing effects of treatments. With magnetic resonance imaging (MRI) now assisting diagnosis and early detection of MS, there is an even greater need for sensitive measures of disease progression to monitor treatment effects at all stages and understand disease evolution.

There is increasing evidence that damage to myelin, axons and neurons result in brain atrophy. Although brain atrophy provides a measure of one aspect of disease progression in MS, highly sensitive, reproducible, robust and precise techniques are required in order to detect subtle treatment effects. A number of manual, semi-automated and automated techniques have been developed to measure rates of brain atrophy from serially acquired MRI scans. It is unclear however which of these different methodologies provides the most effective measure of atrophy rate in MS, and therefore which could be utilised most efficiently as a marker to monitor the disease process.

Sample sizes needed to power therapeutic trials are driven by the variance in outcome measure. Even a small reduction in the variance of atrophy rates could have a major impact on the power of studies in MS, potentially allowing smaller and more cost effective trials to be conducted. It would in turn allow effective treatments to be provided more rapidly to people with MS in addition to reducing the number of trial subjects exposed to ineffective treatments, or treatments with significant side effects. This thesis aims to optimise, assess and compare different brain atrophy measurement techniques in terms of their ability to quantify rates of cerebral atrophy for use as a marker of disease progression in MS.

The project will assess the association of brain atrophy rate with clinical measures, with a view to aid in our understanding of the disease process. Clinical correlation is also essential to establish the meaningfulness and relevance of an MRI measure of disease

progression. Development of new and existing research tools used in the analysis of MRI may help us to determine the distribution and rate of disease activity. A greater understanding of the early phase of MS may also aid in the development of new disease-modifying drugs.

Aims

To investigate a wide range of image analysis techniques of potential value in the measurement of disease progression and assessment of therapeutic efficacy in MS by:

1. comparing and cross-validating different MRI methodologies for longitudinal analysis of brain atrophy, including current methods used in MS research, together with novel techniques;
2. estimating the number of subjects necessary to detect a disease-modifying effect in a trial of a putative disease-modifying drug, using brain atrophy rate as a marker of disease progression;
3. using MRI techniques to understand the longitudinal pattern of brain atrophy in relapsing remitting MS and its relationship to clinical disease progression.

1 Neuroimaging in multiple sclerosis

1.1 Introduction

Multiple sclerosis (MS) is a chronic inflammatory autoimmune disease of the central nervous system (CNS) and the most common neurological condition affecting young adults, with approximately 100-200 cases per 100,000 persons in the United Kingdom (Forbes & Swingler, 1999). Females are more often affected than males by a ratio of 2:1 (Mumford *et al*., 1992), and age at clinical onset is typically between 20 and 40 years of age (Williams & McKeran, 1986). In about 85% of cases, the early course is with relapses and remissions (relapsing remitting MS (RRMS)), while in 15% there is steadily progressive disability from onset (primary progressive MS (PPMS)). After a period that is usually several years, or sometimes decades, progressive disability ensues in about two thirds of those with a relapsing-remitting onset (secondary progressive MS (SPMS)). Symptoms vary between patients but include muscle weakness, problems with balance, loss of coordination and mobility, visual and sensory problems, difficulties with speech, bowel, bladder and sexual dysfunction and mild impairments in cognition. The disease therefore often causes a significant personal, social and economic impact for patients and healthcare services.

Magnetic resonance imaging (MRI) exploits the properties of hydrogen molecules within different tissues and fluid, allowing pathological changes within the brain and spinal cord to be monitored *in vivo*. It is now established as a key investigation in the diagnosis of MS and is increasingly used in studies seeking to monitor disease progression. It may also help us understand the underlying mechanisms and course of MS. MRI can easily be applied at the earliest stages of disease, when samples for histopathological study are unlikely to be available, allowing indirect assessment of pathology. It is a non-invasive technique and unlike computerised tomography does not utilise ionising radiation, making it more practical for repeated examination of people with a condition that may last decades.

Quantitative MRI has confirmed - and to an extent preceded - post-mortem studies showing widespread abnormalities in the brain well beyond obvious “lesions”. Furthermore these studies provide evidence for the important roles of axonal damage and

neurodegeneration in the pathogenesis of MS. Conventional MRI (e.g. T2-weighted) has allowed focal demyelinating lesions within the white matter (WM) and more rarely grey matter (GM) to be visualised and quantified in both post-mortem tissue and *in vivo* (Bø *et al*., 2007; Calabrese *et al*., 2007b; Geurts *et al*., 2005; Molyneux *et al*., 1998c). Whilst these areas of focal WM damage have been studied extensively, the knowledge that pathological changes are occurring in normal appearing tissue has increased support for the measurement of tissue loss (atrophy) from MRI as a marker of overall tissue damage and neuroaxonal loss. A number of MRI-based techniques measuring CNS atrophy *in vivo* have shown progressive brain atrophy, at a rate greater than that seen in normal aging, in subjects with MS (Chard *et al*., 2004; Rovaris *et al*., 2005a); this has been associated with disability progression over long-term studies (Fisher *et al*., 2000).

This chapter will begin by considering the disease of MS before focussing on MRI measures of brain atrophy in MS. However for inclusiveness and to allow comparison with brain atrophy measures, it will firstly address conventional MRI of lesions and quantitative MRI techniques that have been used increasingly in the last 10 years. A discussion of some of the practical issues involved, methods developed for measuring atrophy, and the application of atrophy measurements in clinical and research environments will be presented. Whilst important studies have documented MRI changes in the spinal cord and optic nerves in patients with MS, this thesis will focus only on brain MRI.

1.2 Multiple sclerosis

1.2.1 Pathogenesis of multiple sclerosis

Whilst the exact causes of MS are unknown, several genetic and environmental risk factors have been suggested, and MS is likely to occur as a complex combination of these, and as yet unidentified, influences. Postulated risk factors include the alleles associated with major histocompatibility complex molecules, infectious agents such as the Epstein-Barr virus and Chlamydia pneumoniae, lack of exposure to sunlight and vitamin D, and smoking (Levin *et al*., 2005; Lincoln *et al*., 2005; Pekmezovic *et al*., 2006; van der Mei *et al*., 2003; Yao *et al*., 2001).

Although inflammatory WM lesions in the CNS are the hallmark of the disease, the pathology is now understood to be more extensive, with areas of focal demyelination

occurring in the GM (Bø *et al*., 2003b; Geurts *et al*., 2005; Kidd *et al*., 1999; Peterson *et al*., 2001), gliosis within lesions (van Walderveen *et al*., 1998), diffuse inflammation (Kutzelnigg *et al*., 2005) and progressive neuroaxonal loss in normal appearing tissue (Bjartmar *et al*., 2001; Evangelou *et al*., 2000a). Microscopic examination of MS post-mortem brain tissue has provided direct evidence of neuroaxonal damage including a decrease in neuronal size, disturbance of the neuronal cell cycle, neuronal death, axonal transection, neuronal loss, and dendritic and synaptic loss (Bitsch *et al*., 2000; Lu *et al*., 2000; Peterson *et al*., 2001; Trapp *et al*., 1998; van Waesberghe *et al*., 1999). Decreases in axonal number and density have been shown within chronic and acute WM lesions (van Waesberghe *et al*., 1999; van Walderveen *et al*., 1998) and normal appearing WM (NAWM) (Evangelou *et al*., 2000b), and axonal transection and loss underlies GM lesions found at post-mortem (Peterson *et al*., 2001). Axonal spheroids, transections and abnormal constrictions and dilatations have all been shown both within lesions and their surrounding tissue using immunostaining (Kuhlmann *et al*., 2002; Trapp *et al*., 1998). Whilst inflammation and demyelination are reversible to an extent, neuroaxonal damage is permanent and is likely to be a relevant mechanism of permanent disability.

The mechanisms by which MS pathology occurs are clearly complex and not entirely understood at present. It is likely that different mechanisms operate in different patients and that a number of processes overlap, for example demyelination and remyelination. However evidence suggests that initially, inflammatory cells cross into the CNS through the blood brain barrier, the integrity of which is altered in patients with MS particularly during the acute stages of the disease (Leech *et al*., 2007; Soon *et al*., 2007). The immune response involving T-lymphocytes, macrophages and microglia is propagated by the expression of major histocompatibility complex molecules and release of cytokines. Ultimately there is destruction of oligodendrocytes and myelin (Barnett & Prineas, 2004; Bitsch *et al*., 2000) most likely involving cytokines, glutamate, macrophages, reactive oxygen species and proteolytic enzymes. Demyelination of the GM appears to occur independently of WM change (Bø *et al*., 2007) and involves less inflammation (Bø *et al*., 2003a), but the exact mechanisms are unclear.

Evidence of neuroaxonal degeneration within lesions, particularly active lesions, supports the idea that this pathology may be a consequence of acute focal inflammation (Peterson *et al*., 2001; Trapp *et al*., 1998). Numbers of $CD8^+$ T-cells increase in inflammatory

lesions (Babbe *et al*., 2000) and myelin breakdown leaves axons vulnerable to direct attack from inflammatory and other biological mediators such as proteolytic enzymes, soluble antibodies, cytokines, glutamate, oxidative products and free radicals. A positive correlation of $CD8^+$ T-cells, macrophages and microglia with the extent of axonal damage has been shown (Bitsch *et al*., 2000; Kuhlmann *et al*., 2002). In particular, nitric oxide has been identified as contributing to neuroaxonal degeneration within lesions through changes which include altering mitochondrial DNA and energy metabolism; this mechanism may be more prominent during high axonal firing, when energy demands are greater, and leads to calcium-mediated cell death (Dutta *et al*., 2006; Lu *et al*., 2000; Smith *et al*., 2001a). It has also been suggested that defects in astrocyte beta-adrenoceptors cause a decrease in the energy supply to axons (De Keyser *et al*., 2004), and that voltage-gated calcium channels are redistributed following initial axonal injury (Kornek *et al*., 2001), both of which may also lead to calcium-mediated cell death. In addition to exposing axons to direct attack, secondary effects may result from myelin loss. Myelin breakdown products may contribute to the disease process by activating microglia which mediate neuronal damage (Diestel *et al*., 2003). There is also evidence that oligodendrocytes provide trophic support to axons, the loss of which leads to degeneration (Lappe-Siefke *et al*., 2003; Sanchez *et al*., 1996; Wilkins *et al*., 2003). These factors appear to be more important than myelin itself, as degeneration has been shown to occur despite intact surrounding myelin and a lack of inflammation (Bjartmar *et al*., 2001; Lappe-Siefke *et al*., 2003). Lastly, correlations observed between focal lesions and NAWM changes are consistent with Wallerian degeneration whereby axonal transection within lesions causes distant neuroaxonal damage, possibly due to the loss of pre- and post-synaptic signals (Bjartmar *et al*., 2001; Evangelou *et al*., 2000b). Axonal loss within NAWM has been found to correlate with the regional lesion load (Evangelou *et al*., 2000b). There may also be other less well understood but quantitatively important causes of axonal loss.

1.2.2 Clinical onset and diagnosis

Clinical onset of MS is varied, but involves focal or multi-focal neurological symptoms resulting from inflammatory lesions, demyelination and progressive neuroaxonal loss causing disruption in nerve signalling pathways. The heterogeneous location of pathology leads to the diverse symptoms observed but the most common first symptoms reported include changes in sensation in the arms, legs or face, optic neuritis, weakness, double

vision and balance problems. Approximately 15% of patients present with multi-focal symptoms whilst others may have more unusual presentations such as aphasia or psychosis. Whilst in the majority of patients these symptoms will abate over the course of a few weeks, in some patients there is insidious progression of neurological symptoms from onset.

A clinical diagnosis cannot be made by a single test and at initial clinical presentation of symptoms, as evidence of dissemination in time and space of lesions is required. A further clinical attack implicating a different lesion site will allow diagnosis based on objective clinical evidence of two or more lesions. The most recent criteria for the diagnosis of MS, the McDonald criteria (McDonald *et al*., 2001) which were revised in 2005 (Polman *et al*., 2005), incorporate MRI and laboratory tests into the diagnostic scheme. These investigations can provide evidence of dissemination of lesions in time and space, allowing a diagnosis of MS to be given in patients who have experienced only a single clinical monosymptomatic event (so called clinically isolated syndrome (CIS)). Approximately 70% of people who present with a CIS will be subsequently diagnosed with MS (Brex *et al*., 2002) and it has been shown that the McDonald criteria predict a clinically definite MS diagnosis (i.e. a second clinical relapse) in those people presenting with a CIS (Dalton *et al*., 2002b). In patients with an insidious progression from onset, at least one year of disease progression and additional evidence from MRI, cerebrospinal fluid (CSF) and/or visual evoked potentials is required for diagnosis.

1.2.3 Disease course and clinical subtypes

The complex pathogenesis and pathology of MS (both destructive and restorative), pathology occurring in clinically silent locations and cortical plasticity combine to make the clinical disease course of MS extremely unpredictable. Whilst some people will have a relatively benign course of disease others are more likely to develop disability (Sayao *et al*., 2007). In spite of this, several clinical patterns have been characterised in MS (Lublin & Reingold, 1996). Around 85% of patients initially have a relapsing remitting disease type. RRMS consists of clearly defined disease relapses (attacks of acute neurological symptoms) with full or partial recovery and no further progression of disease between relapses. Inflammation and lesion formation are likely to be the precursor to relapses. Of those patients with RRMS most will go on to develop a progressive form of the disease within an average of 20 years (Vukusic & Confavreux, 2003). This is called SPMS and

these patients may have occasional superimposed relapses, minor remissions and plateaus during the progressive phase. PPMS describes the 15% of people who have a progressive form of the disease from onset with gradual but almost continuous worsening of disability and only occasional plateaus and temporary minor improvements in function. New inflammatory lesions are seen less in the progressive stages of the disease. A progressive-relapsing disease course has also been described and is seen in a minority of patients. It is characterised by progressive disease from onset, but with clear acute relapses and continued progression between relapses.

Prognosis for an individual with MS appears to depend to some extent on the early clinical course, gender, age at onset and initial symptoms (Langer-Gould *et al*., 2006). However, around 15-25% of people diagnosed with MS appear not to progress, even after periods of 15 years or more, and are classed as having benign MS (Pittock *et al*., 2004).

Whilst there is no cure for MS several immunomodulatory treatments have been developed and licensed for use in RRMS; these have been shown to reduce the relapse rate and reduce the accumulation of disability (Jacobs *et al*., 1996; Johnson *et al*., 1998; Polman *et al*., 2006; The IFNB Multiple Sclerosis Study Group, 1995). Additionally, one of these treatments, interferon beta-1b, has also been shown to delay the conversion to clinically definite MS and the development of disability in patients presenting with a CIS (Kappos *et al*., 2007). Steroid treatment can be given in the event of a relapse to reduce inflammatory events and expedite recovery from a relapse, but this does not appear to alter long-term prognosis. Symptomatic treatments are available for a number of disease manifestations including tremor, spasticity and incontinence. There is no effective disease-modifying treatment for progressive forms of MS, and current research is focussed on developing neuroprotective therapies to prevent the ongoing neuroaxonal loss thought to underlie progressive disability.

Disease progression is traditionally assessed by neurological examination and disability rating scales. The scale most commonly used is the Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) (Appendix 1), which consists of scoring eight different functional systems. However the EDSS is weighted towards assessment of the motor symptoms of MS and does not consider symptoms such as pain, or the cognitive dysfunction which has been estimated to occur in 30-70% of cases (Amato *et al*., 2001;

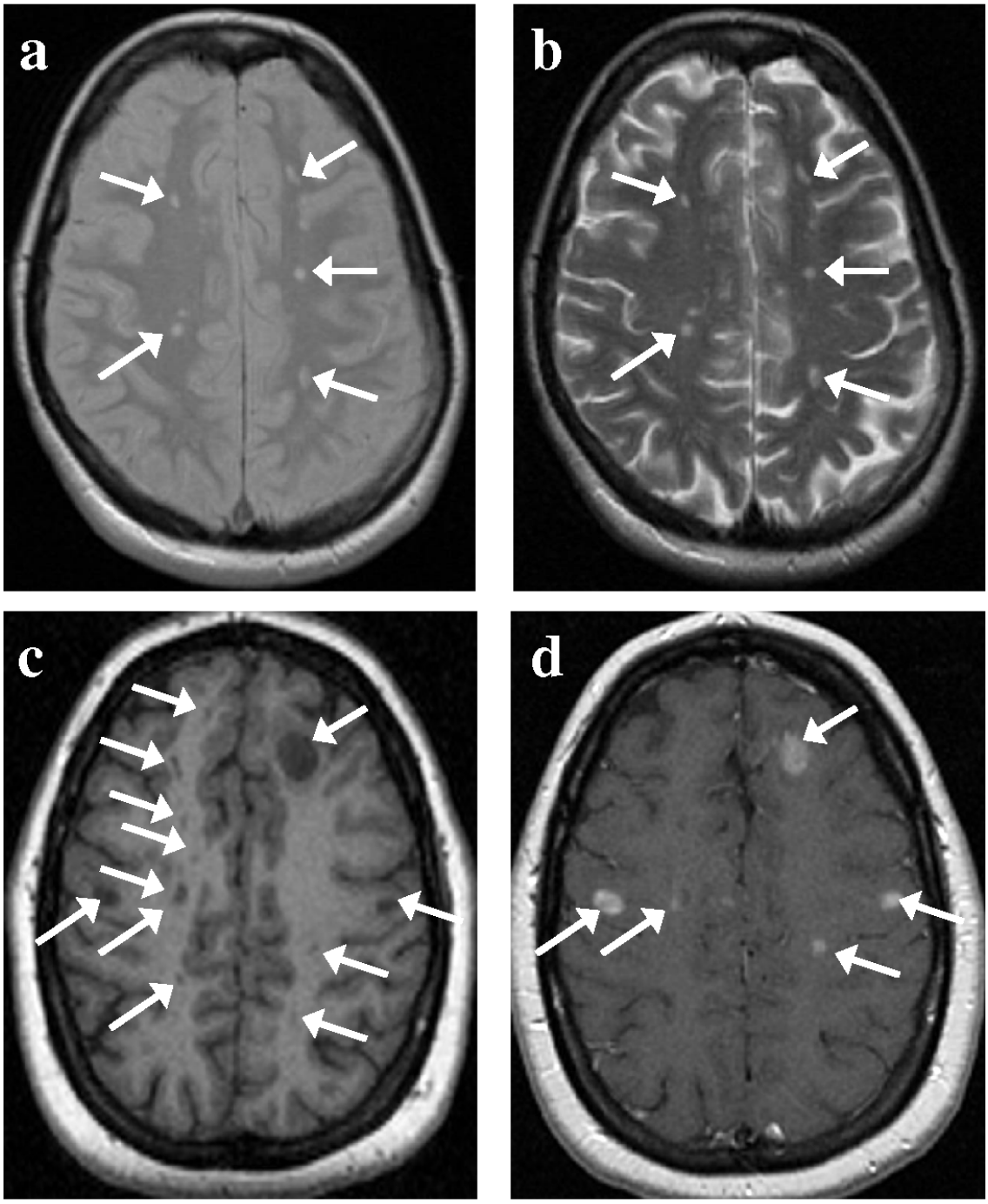
Lazeron *et al*., 2005; Möller *et al*., 1994; Portaccio *et al*., 2006; Rovaris *et al*., 1998). Moreover, at least a one-point sustained change on the EDSS is needed to be confident of progression in the degree of disability, due to the difficulty in precisely defining the level of impairment in some functional categories and inter-rater variability (Noseworthy, 1994). MRI has therefore gained importance in the last 15 years, not only for its role in diagnosing MS, but as a tool for monitoring disease progression in a more objective and potentially more sensitive manner.

1.3 Conventional MRI of focal lesions

Conventional MRI allows visualisation of the structure of the brain. Areas of focal damage within the WM and GM can be visualised *in vivo* due to the increased water content of lesions relative to the surrounding tissues. Conventional spin echo (CSE) and fast spin echo (FSE) sequences, which result in a proton density-weighted image (PD-weighted) and a T2-weighted image, have commonly been used in MS for diagnostic purposes. New and enlarging lesions appear hyperintense on PD- and T2-weighted imaging (Figure 1-1a and Figure 1-1b), although it can be difficult to detect lesions situated near the periventricular border on T2-weighted images, due to the similar contrast of CSF. Fluid attenuated inversion recovery (FLAIR) sequences have partly overcome this problem by suppressing the signal from CSF. Many lesions seen on T2-weighted imaging can also be visualised on T1-weighted images where they appear hypointense (Figure 1-1c) and tend to correspond to areas where there is severe tissue disruption (van Walderveen *et al*., 1998). During the active stage of MS there is increased permeability of the blood brain barrier and the contrast agent gadolinium (Gd) has been used with T1-weighted imaging to identify areas of acute inflammation and distinguish active and inactive lesions (Figure 1-1d).

Whilst visual examination of these images can be performed to aid diagnosis, quantitative methods have been developed to assess the extent of lesion burden as a guide to disease severity and progression. At the simplest level assessment of the number of lesions can be performed and scoring systems based on the number and size of lesions have been applied (Thompson *et al*., 1991). However the intra- and inter-observer reliability is low, even when incorporating consensus criteria and training observers (Filippi *et al*., 1995; Molyneux *et al*., 1999).

Figure 1-1 *White matter lesions visualised on MRI. a) PD-weighted, b) T2-weighted, c) T1-weighted, d) T1-weighted with gadolinium enhancement (indicating active inflammatory lesions). PD- and T2-weighted imaging is from a different subject to T1-weighted imaging.*



An alternative approach is to outline lesions on MRI and quantify the volume, and numerous methods have been developed for this. Whilst manual outlining of lesions may be more accurate it is labour-intensive and subject to intra- and inter-observer

reproducibility problems (Filippi *et al*., 1995; Mitchell *et al*., 1996). Semi-automated methods for delineating lesions have been shown to be equivalent to manual outlining with regard to accuracy but with improved reproducibility (Filippi *et al*., 1995; Grimaud *et al*., 1996; Molyneux *et al*., 1998a). These include global and local thresholding (Grimaud *et al*., 1996; Molyneux *et al*., 1998c; Wicks *et al*., 1992), and cluster-based identification and delineation of lesions based on fuzzy-connectedness principles (Udupa *et al*., 1997). Unfortunately semi-automated methods may require almost the same level of operator input as manual delineation (Grimaud *et al*., 1996). Manual identification of lesions prior to contouring may be needed, or errors in lesion identification (false negatives or false positives) and incorrect contouring using a global threshold may need correcting. Even when voxel intensities are standardised across an image, sensitivity may still be poor (Molyneux *et al*., 1998c).

Developments in the area of fully automated lesion identification and delineation have mainly involved the use of multispectral image data and studies of these algorithms have shown that the accuracy, specificity and reproducibility of lesion volumes are comparable to manual or local threshold methods (Achiron *et al*., 2002; Datta *et al*., 2006; Wu *et al*., 2006). Automated techniques have also been useful in the longitudinal assessment of lesion load. New, enlarging and shrinking lesions can be more easily detected following rigid or non-rigid registration of images and can be used to distinguish between gain and loss in lesion volume rather than net change only (Rey *et al*., 2002; Tan *et al*., 2002a). Registration overcomes problems associated with repositioning errors that may lead to bias when only small changes in lesion volume have occurred (Gawne-Cain *et al*., 1996) and automated analysis has been shown to identify a greater number of evolving lesions than manual detection (Bosc *et al*., 2003).

Despite lesions being an obvious pathological feature of MS, their correlation with disability and disease progression has been mixed (Dastidar *et al*., 1999; Held *et al*., 2005; Kappos *et al*., 1999; Sailer *et al*., 2001; Stevenson *et al*., 2004). In one study, correlation between the change in lesion volume over 0-5, 5-10 and 10-14 years, and EDSS score at year 14 ranged between 0.29 and 0.61 (Brex *et al*., 2002). This is particularly true in patients with PPMS in whom a smaller MRI lesion load and less lesion activity is seen than in patients with RRMS and SPMS, but disability progresses nonetheless (Revesz *et al*., 1994; Thompson *et al*., 1991). Several reasons may lie behind this paradox. Firstly,

lesions may be heterogeneous, both in the extent of demyelination, remyelination and permanent tissue disruption that has occurred (Lucchinetti *et al*., 2000), and in the location they develop. Both these factors are likely to influence the effects that lesions have on the extent of permanent disability (Charil *et al*., 2003). Secondly, monitoring lesions does not encompass the global changes that are now known to occur in normal appearing tissue. Thirdly, the MRI acquisition will affect both detection of lesions and lesion measurements. Typically, lesion measurements are performed on MRI with slices that are a minimum of 3mm thick, and this may cause small lesions to go undetected because they are smaller than the resolution of the scan. In addition measurements from these acquisitions will be more prone to repositioning errors over serial studies and the effects from partial volume voxels at lesion edges will be greater. Although volumetric FSE and FLAIR acquisitions exist (with slice thicknesses as little as 1mm), and have been shown to increase the number of detectable lesions compared with conventional acquisitions (Ciccarelli *et al*., 2002; Molyneux *et al*., 1998b; Tan *et al*., 2002b), the time required for analysis using manual or semi-automated techniques will be greater and correlations with disability may not improve (Ciccarelli *et al*., 2002). It has also been shown that lesions occur within the GM and at the GM/WM boundary, but often go undetected on MRI using conventional sequences at the field strengths currently used at the majority of centres (1.5 Tesla (T) and 3T) (Geurts *et al*., 2005; Kangarlu *et al*., 2007).

1.4 Non-conventional MRI to assess global effects in multiple sclerosis

Non-conventional quantitative MRI techniques have been developed which have provided indirect evidence of the global pathology occurring in MS and aided in our understanding of the disease. More sophisticated methods are required for image acquisition and image post-processing than conventional MRI and these techniques are therefore mainly limited to research settings.

1.4.1 Diffusion imaging

Diffusion of water molecules inside biological tissues can be measured *in vivo* using diffusion weighted MRI (DW-MRI). The motion of water molecules is hindered by microstructural barriers including cell membranes and organelles, meaning that diffusion is lower in brain tissue than in free water. Pathological processes in MS may change the structural barriers within the brain, increasing permeability and diffusivity and providing information regarding tissue integrity. An extension to DW-MRI is diffusion tensor

imaging (DTI) which allows the motion of water molecules in all directions to be characterised. WM has an organised structure of aligned axons and diffusion is greater along the axis of axonal fibres compared with across the fibres; this feature of diffusion (the property of anisotropy) can provide additional information regarding changes in tissue structure. In addition, DTI has been used to perform tractography and determine WM connectivity (Pagani *et al*., 2005a). Whilst early DW-MRI studies were limited to region of interest (ROI) analysis, which is subject to selection bias, reproducibility problems and insensitivity to overall disease effect, methodology has been developed that has allowed analysis of global changes (Cercignani *et al*., 2001b). However the origin of any changes cannot be determined and partial volume effects need correcting for, particularly when atrophy is occurring (diffusion will be greater in partial CSF voxels than pure GM or WM voxels) (Rashid *et al*., 2004).

Despite these limitations, DW-MRI studies have shown that diffusion in both MRI-visible lesions and NAWM of MS patients is significantly greater than in controls, suggesting an increase in permeability due to myelin damage and axonal loss (Coombs *et al*., 2004; Droogan *et al*., 1999; Oh *et al*., 2004a; Vrenken *et al*., 2006b). In addition, anisotropy within lesions and NAWM is decreased in patients with MS, suggesting there is disruption to the structural organisation of tissue (Droogan *et al*., 1999; Oh *et al*., 2004a; Vrenken *et al*., 2006b). Increased diffusivity has also been shown within the normal appearing GM (NAGM) and basal ganglia of patients with MS, providing further evidence that MS pathology is not merely restricted to the WM (Bozzali *et al*., 2002; Cercignani *et al*., 2001a). Moreover, increased diffusivity has been demonstrated in people presenting with a CIS, suggesting that these changes are present from the earliest stages of disease (Gallo *et al*., 2005; Ranjeva *et al*., 2003). However, these measurements are altered by temporary damage to tissues, such as demyelination, and it is difficult to determine whether changes in diffusion and anisotropy are reversible or permanent from cross-sectional studies. Longitudinal studies however have demonstrated progressive changes in diffusivity in NAWM and NAGM in patients with CIS, RRMS and progressive MS (Caramia *et al*., 2002; Cassol *et al*., 2004; Garaci *et al*., 2007; Oreja-Guevara *et al*., 2005; Rovaris *et al*., 2005a).

One of the current limitations of DW-MRI is the trade-off between long acquisition times and spatial resolution. Low spatial resolution can lead to reproducibility problems for

ROI analysis and increase partial volume effects in global analyses. However, techniques are being developed that could provide high-resolution 3D acquisitions with minimal increases in acquisition time (Cercignani *et al*., 2005). DW-MRI also requires standardisation of hardware and acquisition protocols in order for findings to be comparable across patients.

1.4.2 Magnetisation transfer imaging

Magnetisation transfer MRI (MT-MRI) quantifies the ratio between the concentration of protons in the brain that are free, for example as tissue water, to those that are bound up as macromolecules in myelin and other cell membranes and therefore restricted in motion. The MT ratio (MTR) between free and restricted protons can be measured. A change in the organisation of brain tissue as a result of demyelination and neuroaxonal damage reduces the number of protons bound up as macromolecules and MTR decreases as a result. In post-mortem tissue MTR has been correlated with axonal density and myelin content (Schmierer *et al*., 2004; van Waesberghe *et al*., 1999). As with DW-MRI, MT-MRI can be conducted on an ROI or on a global level through histogram analysis of MTR values over the whole brain, NAWM or NAGM.

In vivo MT-MRI has shown a decrease in MTR values in lesions and in normal appearing brain tissue compared with controls in patients presenting with a CIS (Audoin *et al*., 2004; Fernando *et al*., 2005; Rovaris *et al*., 2003) and MS (Davies *et al*., 2004; Traboulsee *et al*., 2003; Vrenken *et al*., 2006a). Although MT-MRI may be a good way to help elucidate pathological mechanisms, it is not a completely specific marker of demyelination and neuroaxonal damage. Inflammation alone could cause MT values to fall, as the density of macromolecules is diluted by oedema. There may be partial or complete recovery of MTR over a period of months as inflammation subsides and remyelination occurs. However longitudinal studies have shown that there is a progressive decrease in MTR values over one year in patients with a CIS and MS (Rovaris *et al*., 2003) and these have been shown to correlate with worsening of disability over an eight year period (Agosta *et al*., 2006). Voxel-wise statistical analyses of global MTR values have been performed in CIS and RRMS cohorts, which have allowed localisation of changes relative to controls without the *a priori* assumptions involved with ROI analysis (Audoin *et al*., 2004; Audoin *et al*., 2007a).

As with DW-MRI, partial volume effects will alter MTR values, particularly over longitudinal studies when atrophy is occurring, although normalising MTR values for brain volume and applying a strict threshold for voxel inclusion should minimise these effects. Again, standardisation of scanners, acquisition parameters and protocols would need to be performed if MT-MRI were to be used in a wider clinical or research setting.

1.4.3 Proton magnetic resonance spectroscopy

Proton magnetic resonance spectroscopy (^{1}H-MRS) allows a number of metabolites common in the brain to be measured. Changes in these metabolites can be used to infer inflammation, demyelination and neuronal damage. The major resonances are i) choline (contained in phospholipids), ii) creatine and phosphocreatine (Cr), iii) N-acetyl groups (mainly N-acetylaspartate (NAA)), iv) lactate. Choline, lactate and Cr are thought to be markers of acute inflammation or demyelination. Membrane lipids containing choline increase during active myelin breakdown and lactate may increase as a result of the increase in metabolism by inflammatory cells. Likewise, decreased levels of NAA, an amino acid found within mature neurons, are thought to provide indirect assessment of neuronal integrity.

Decreased NAA has been observed within normal appearing brain tissue (both WM and GM) in MS subjects compared with controls (Chard *et al*., 2002a; De Stefano *et al*., 2001; Oh *et al*., 2004b; Sastre-Garriga *et al*., 2005b), providing evidence from yet another MR modality that neuroaxonal damage is a consistent feature of MS. Whilst these findings have also been found in studies of CIS (Filippi *et al*., 2003; Rovaris *et al*., 2005b), supporting the hypothesis that neuroaxonal damage occurs from the earliest stages of MS, normal NAA in NAWM has also been noted in these subjects (Fernando *et al*., 2004; Ranjeva *et al*., 2003) and MS patients (Vrenken *et al*., 2005). Normal metabolite concentrations have also been found within the cortex of MS patients (Geurts *et al*., 2006). The application of methods to obtain absolute NAA concentrations, rather than relative to other metabolites, has decreased the possibility of false effects and error in data interpretation. However ^{1}H-MRS may still be limited in detecting the low signal-to-noise ratio metabolite signals when voxel of interest methods are applied, which may lead to a lack of sensitivity to global changes, reproducibility problems and selection bias. Methods for whole brain quantification of NAA have been developed but these do not allow the localisation of changes (Adalsteinsson *et al*., 2003; Gonen *et al*., 2000).

Changes in NAA may not represent neuroaxonal loss, but rather be affected by axonal metabolic function independent of structural integrity (Cader *et al*., 2007). Indeed longitudinal studies have shown recovery or increases in NAA over time (Audoin *et al*., 2007c; Tiberio *et al*., 2006). Single-voxel spectroscopy may be sensitive to changes in hardware and acquisition, which makes implementation on a multicentre scale difficult.

1.5 Brain atrophy as a marker of neuroaxonal damage in multiple sclerosis

Unlike inflammation and demyelination, neuroaxonal degeneration is irreversible, and is likely to result in brain atrophy through the loss of tissue. Its measurement would therefore provide a marker of MS pathology particularly relevant to clinical disability. Furthermore, diffuse neuroaxonal losses across the whole brain can feasibly be tracked. Despite the potential value of brain atrophy as a marker of neuroaxonal loss however, it should be pointed out that a) cerebral volume is composed of many different cell types and may therefore be altered by many changes other than neuronal or axonal loss, and b) other pathological and physiological factors may alter brain volume and should be taken into account when interpreting results (Table 1-1).

1.5.1 Definition of qualities of a good brain atrophy measure

Rating scales for the visual assessment of atrophy have been used in MS, but can be difficult to apply and interpret (Benedict *et al*., 2002; Rao *et al*., 1985). A good quantitative brain atrophy measurement technique would demonstrate many of the features listed in Table 1-2.

Measurement accuracy and precision are desirable, however they are less important in practical terms than sensitivity to disease-related change and measurement reproducibility. Techniques that can be automated and are robust to differences in acquisition are particularly relevant qualities for large longitudinal multicentre studies. In addition, methods that work on volumes (three dimensions) as opposed to on a slice-by-slice basis may be advantageous, as they should be more accurate due to higher spatial resolution.

Table 1-1 *Pathological and physiological variables that may affect brain volume measurements in MS.*

Factors causing brain volume increases	***Factors causing brain volume decreases***
Oedema	Axonal loss
Inflammation	Neuronal damage with neurone loss (e.g.
Gliosis (tissue bulk)	dentritic pruning)
Remyelination	Resolution of inflammation and oedema
	Gliosis (retraction scarring)
	Demyelination
	Dehydration
	Anti-inflammatory agents
	Normal aging

Table 1-2 *Desirable qualities of a brain atrophy measure.*

Quality	***Comment***
Sensitive to brain atrophy	Allows subtle pathological changes to be detected. Techniques that can measure differences at a subvoxel level may be advantageous.
Reproducible	Avoids measurement errors that may lead to erroneous results.
Accurate	Detects actual tissue loss. Accuracy is difficult to verify however, and small errors are insignificant if consistent between subjects and over time.
Precise	Repeated measurements of the same volume will be of the same value to within a small percentage of the volume.
Automated	Fast to implement, thereby reducing operator time and costs. Operator-dependent errors are minimised.
Robust to image quality	Results are more reliable and comparable between subjects and imaging sites where acquisitions may vary slightly, and minimally affected by image artefacts.

The reproducibility and precision of atrophy measurement techniques depend to some extent on variations in the scanner and imaging parameters (i.e. echo time, repetition time, flip angle, slice thickness etc); greater precision in brain parenchymal fraction (BPF) has been obtained on dual-echo T2-weighted and FLAIR images compared with T1-weighted volumetric images from the same subjects (Horsfield *et al*., 2003). In addition fluctuations in the performance of scanner gradients can lead to drift in voxel sizes (Freeborough *et al*., 1996). Although this variation may at least partly be corrected for by normalisation to a constant such as skull, brain atrophy measures that are less affected by gradient strengths are preferable. Reproducibility and precision may also be affected by the quality of images and presence of artefacts.

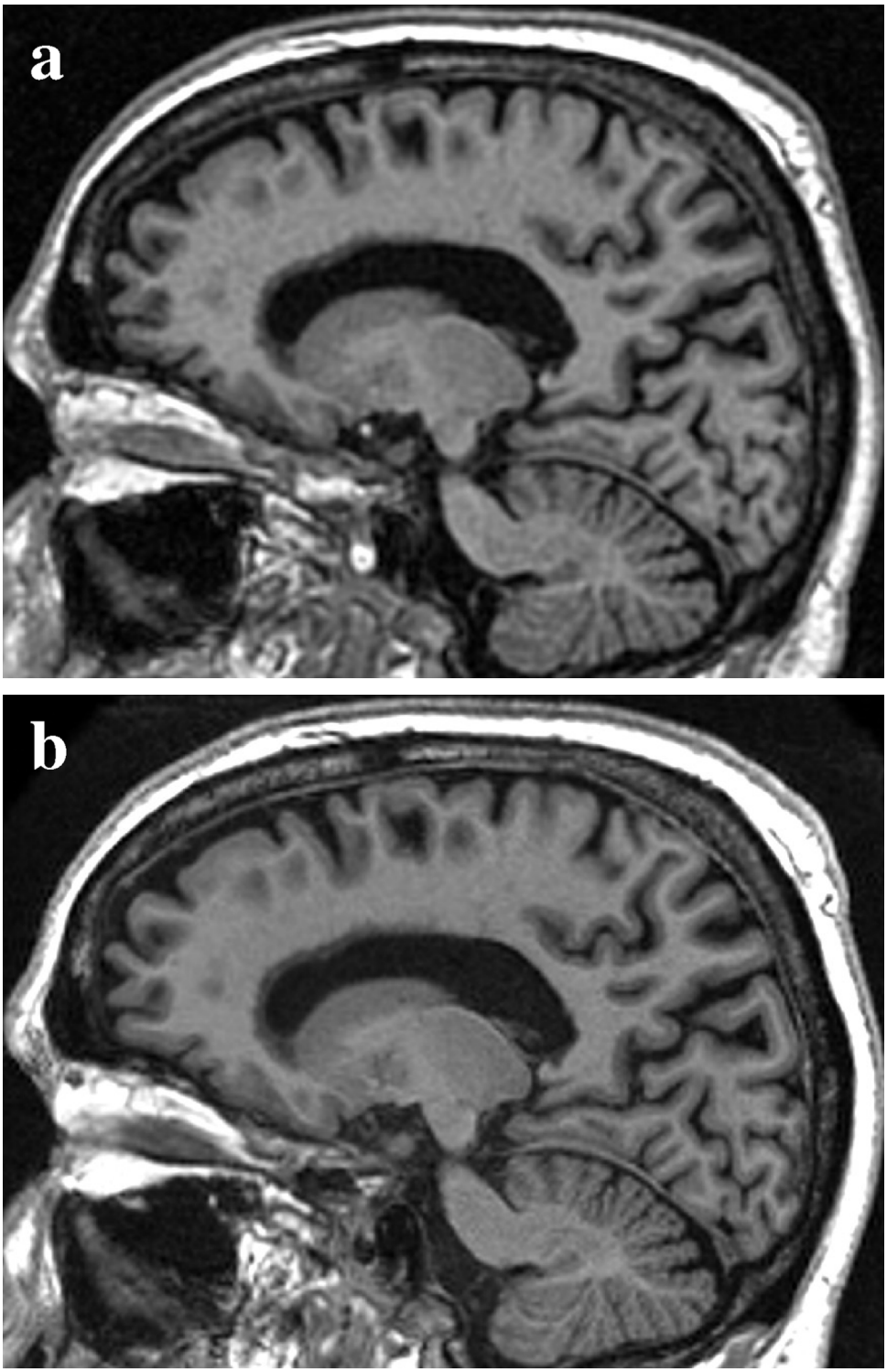
1.5.2 MR acquisition for brain atrophy measurement

The optimum MRI acquisition for atrophy quantification may to some extent depend on the measurement technique and the subjects to be studied. However it would be expected that the optimal acquisition would be one with a high signal-to-noise ratio (SNR), high contrast-to-noise ratio (CNR), high spatial resolution and short acquisition time.

Signal-to-noise ratio

SNR can be defined as the ratio of the mean voxel signal (from a homogenous region with high signal intensity within the object of interest) divided by the standard deviation of the background signal (measured from several regions outside the object). Alternatively a difference image can be generated from two consecutively acquired images, and the mean voxel signal within this image divided by the standard deviation of voxels in this same region. Increasing MR scanner field strength will amplify the signal intensity from the object and consequently SNR (Figure 1-2). Most MR scanners currently in use operate at 1.5T or 3T, although in research settings higher magnetic fields are sometimes used. SNR can also be improved by imaging larger voxels which can be achieved by increasing the field of view (FOV) whilst maintaining the matrix size, and changing the radio frequency receiver coil to a phased array (multi-channel) coil. Atrophy measures will benefit from increased SNR which “sharpens” the edges of tissue boundaries.

Figure 1-2 *T1-weighted images showing different signal-to-noise ratios due to scanner field strength. Images are from a single subject and corrected for intensity inhomogeneity. Signal-to-noise ratio is lower when acquired on (a) a 1.5 Tesla scanner (9:36 minute acquisition time) than on (b) a 3 Tesla scanner (9:14 minute acquisition time). Images are from the Alzheimer's Disease Neuroimaging Initiative (*http://www.loni.ucla.edu/ADNI*).*



Contrast-to-noise ratio

The CNR can be defined as the ratio of the difference in signal intensity between regions with different cellular constituents, and the background signal. The contrast of images must be good to obtain robust brain atrophy measurements as many techniques rely on high contrast boundaries between brain and CSF (Figure 1-3). Regional atrophy measures are also likely to benefit from a high contrast between GM and WM. Good tissue contrast requires selection of an appropriate pulse sequence (T1-weighted sequences are often used), and increasing field strength will also increase CNR.

Spatial resolution

Spatial resolution is the distance between adjacent voxels in the image and therefore to achieve higher resolution and visualisation of more detail, image voxel size must be reduced (Figure 1-4). Although atrophy measurement techniques have been successfully applied to "2D" acquisitions with slice thicknesses of 3mm or greater (Collins *et al*., 2001; Losseff *et al*., 1996; Rudick *et al*., 1999), higher resolution imaging should improve atrophy measures. Partial volume effects result when multiple tissue types contribute to a voxel and there is blurring of intensity across boundaries. It is common to acquire 3D volumetric acquisitions with isotropic voxels of around 1mm which reduce partial volume effects on some measurements, provide good brain/CSF and GM/WM contrast, and allow visualisation and measurement of small regional structures. Atrophy measures applied to 3D acquisitions are less dependent on slice positioning and slice selection, and may give more accurate results from automated techniques (Sharma *et al*., 2004). In addition 3D acquisitions allow reformatting and accurate re-slicing of data for registration-based techniques, and provide good boundary definition in all views which is important for atrophy measurement. The disadvantage of increasing resolution through smaller voxels is that SNR is decreased and acquisition time is increased. However standard high-resolution 3D acquisitions (e.g. 1 x 1 x 1.2mm^3 voxels) can be acquired in less than 10 minutes.

Figure 1-3 *T1-weighted images showing differences in contrast-to-noise ratio due to changes in scanner hardware (a has higher CNR than b).*

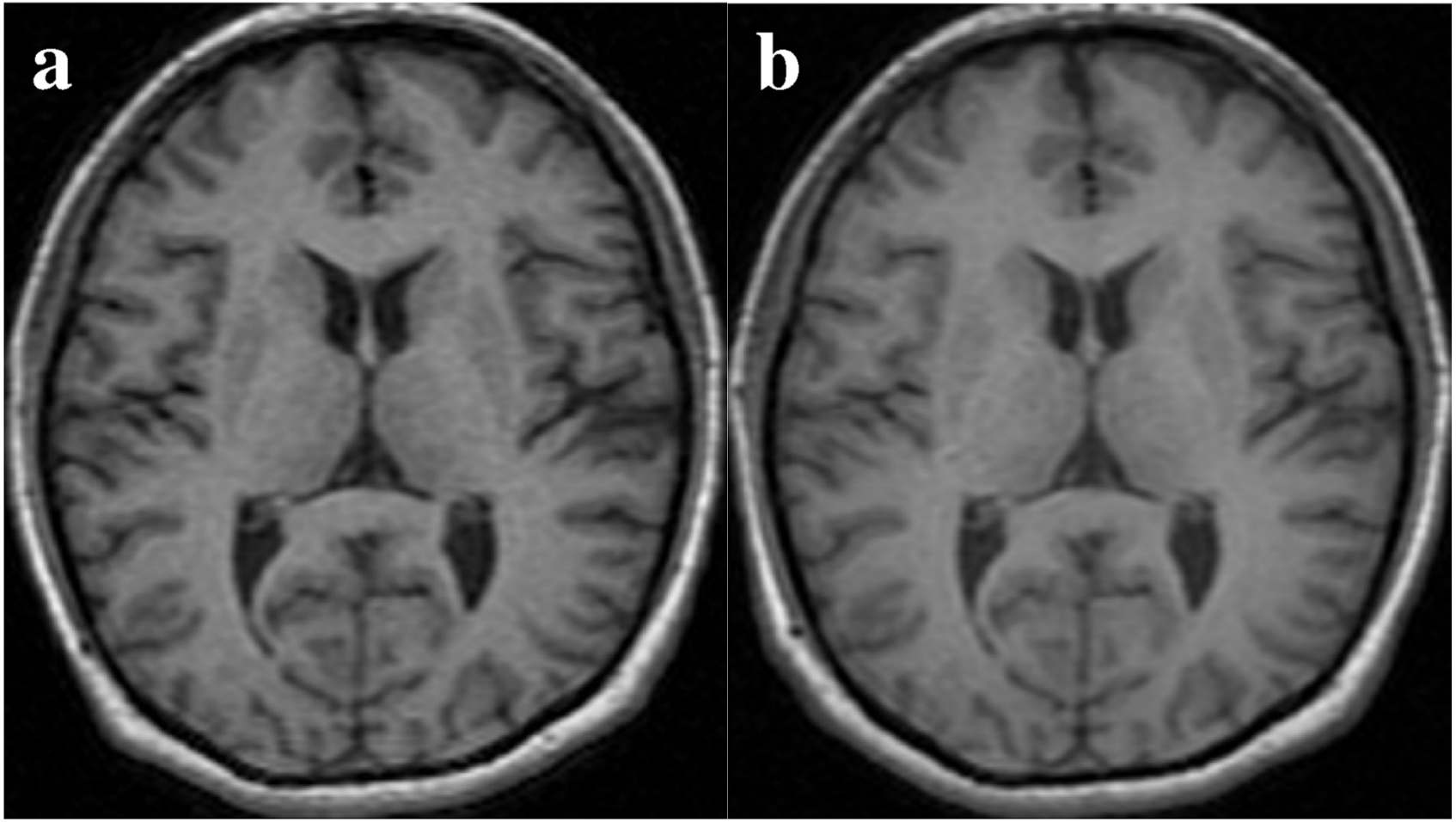
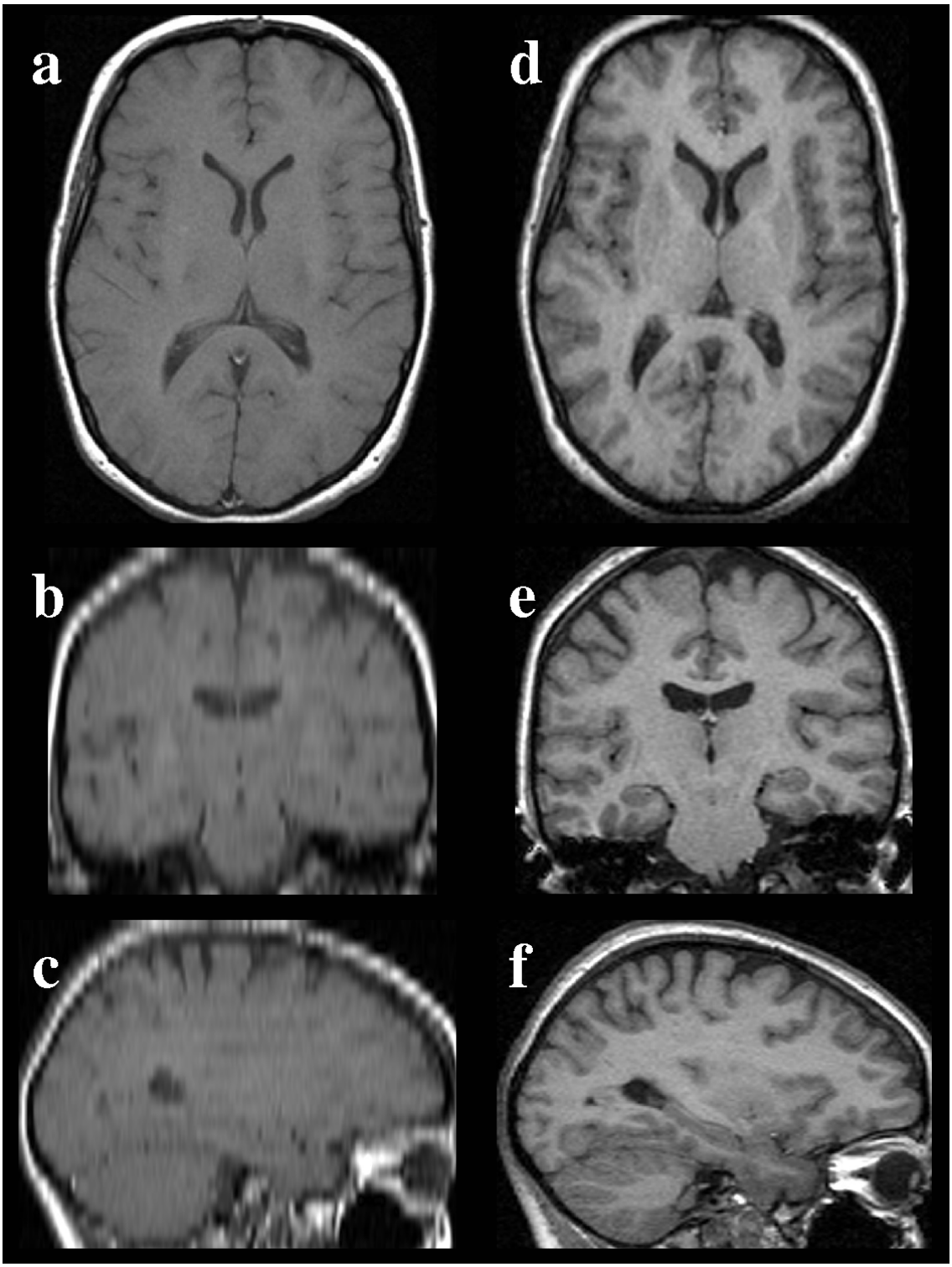


Figure 1-4 *T1-weighted images showing differences in spatial resolution. a)-c) show an image acquired with 3mm thick axial slices (1x1mm in-plane resolution), giving limited resolution in the coronal and sagittal planes (non-isotropic voxels), d)-f) show an image acquired with 1.5mm thick coronal slices (1x1mm in-plane resolution), giving similar resolution in the axial and sagittal planes (near isotropic voxels).*



1.5.3 MRI artefacts

When imaging the brain, MRI artefacts can result from different sources including the scanner, the patient and the MR acquisition. These may only affect a few voxels or they may be on a larger scale affecting visualisation and processing of the image. Some of the more common artefacts that arise and that may affect brain atrophy measures will be addressed in the following section. It should be pointed out that for acquisition-related and patient-related artefacts, the degree of artefact will often be worse the higher the field strength of the scanner. Although the artefact may be present at lower field strengths, the increase in SNR and CNR makes them more conspicuous.

Inhomogeneity

A common artefact seen on MR images of the brain is intensity inhomogeneity (bias), where the signal intensity from supposedly homogeneous tissue (i.e. GM, WM, CSF, skull etc) is non-uniform. These bias fields usually vary smoothly across an image (Figure 1-5a) and may be due to poor radio frequency coil uniformity leading to a non-uniform B1 field (the strength of the radio frequency pulse varies at different positions within the coil), non-uniform sensitivity of the receiver coil and eddy currents caused by magnetic field gradients. Anatomical variability, regional differences in the magnetic properties of the tissues being imaged, position of the head within the MRI scanner and electrodynamic interactions with the object being imaged may also result in variations in the image signal, particularly at higher field strengths.

This inhomogeneity across an image can cause inaccuracies in atrophy measurement techniques that rely on homogeneity of intensity within a tissue class. Techniques that depend on operators recognising tissue boundaries may also be influenced, therefore correcting this intensity inhomogeneity within an image may improve atrophy measurements. Inter-slice variations observed in 2D sequences can be dealt with by methods that normalise the intensities between individual slices. However the smooth intensity variations present in most acquisitions are approached differently. Several post-processing techniques have been developed to retrospectively correct for these inhomogeneities, using a number of different approaches (Ahmed *et al*., 2002; Chen & Reutens, 2005; Cheng & Huang, 2006; Cohen *et al*., 2000; Gispert *et al*., 2004; Luo *et al*., 2005; Van Leemput *et al*., 1999; Vokurka *et al*., 1999; Vovk *et al*., 2004). In addition some segmentation algorithms simultaneously delineate brain regions and estimate and

correct for the inhomogeneity. This procedure offers the advantage that information from the segmentation can be used to aid inhomogeneity correction.

A study published in 2001 directly compared six commonly used algorithms for intensity inhomogeneity correction and found that two methods which evaluate spatial variation in tissue intensity parameters generally performed better with regard to the accuracy, precision and stability of the non-uniformity correction (Arnold *et al*., 2001). These were the bias field corrector (Shattuck *et al*., 2001) and nonparametric nonuniform intensity normalisation (N3) (Sled *et al*., 1998). Indeed, N3 has been shown to increase the reproducibility of brain segmentation in a study of 10 control subjects (Chard *et al*., 2002c).

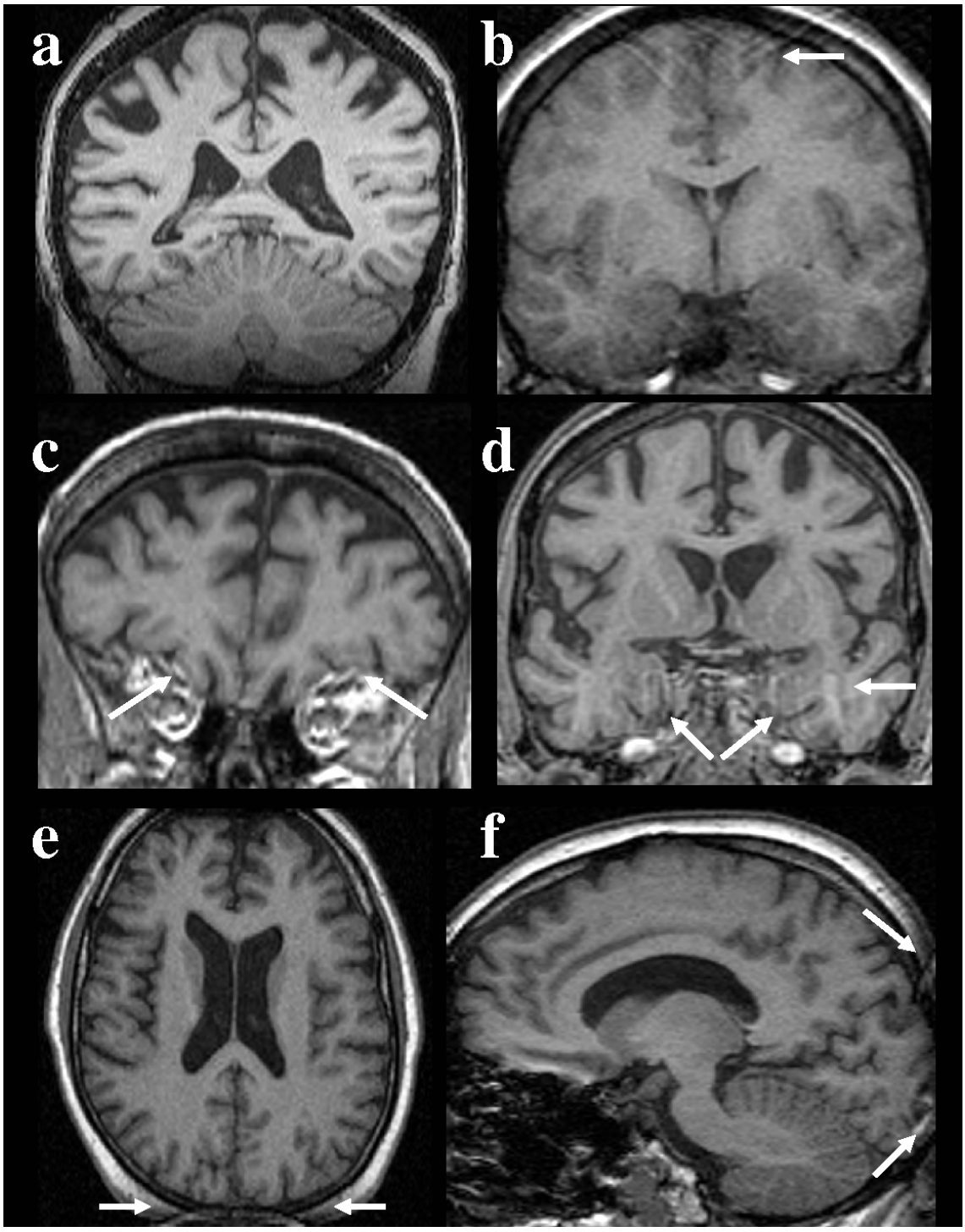
Motion

Motion of the head during scanning can lead to artefacts (Figure 1-5b). Many patients are unable to cooperate in keeping still during scanning and it can be beneficial to avoid long sessions in the scanner when patients may become uncomfortable. Motion effects are also minimised to some extent by the use of head restraints. Motion artefacts may also present as a result of pulsatile blood flow or breathing. These effects can be averaged out of the image or reduced by synchronising the imaging sequence with the cardiac cycle or suppressing the signal from blood. PROPELLER (Periodically Rotated Overlapping ParallEL Lines with Enhanced Reconstruction) MRI has been described as a means by which to quantify and correct for motion artefact prior to image reconstruction (Forbes *et al*., 2001; Pipe, 1999). Movements of the eye and swallowing may also cause motion artefacts and are difficult to correct for.

Chemical shift

Chemical shift describes the artefact that occurs due to differences in the resonance frequencies of fat and water, which cause displacement in the signal from fat relative to water. This displacement can make it difficult to determine brain boundaries (Figure 1-5c). Fat suppression techniques and changes to the imaging sequence can help reduce chemical shift artefacts.

Figure 1-5 *Examples of MR image artefacts. a) intensity inhomogeneity, b) motion, c) chemical shift d) susceptibility and pulsation artefact, e) and f) infolding and inadequate field of view.*



Susceptibility artefacts

This class of problems refers to a number of artefacts caused by the different magnetic susceptibilities of tissues and materials leading to local non-uniformity of the magnetic

field. This can result in signal dropout, bright spots and spatial distortion in images. Metal implants may also cause these artefacts, e.g. dental plates (Figure 1-5d). These artefacts can be reduced using SE or FSE sequences compared with gradient echo sequences and a high bandwidth and short echo time may also help.

Inadequate field-of-view or number of image slices

If the FOV is too small or the number of slices not adequate to include the whole head, this can lead to wrap-around (aliasing) where one side of the image folds into the opposite side of the image (Figure 1-5e). It is caused by a corruption in the spatial encoding of objects outside the FOV which cannot be distinguished from objects inside the FOV. This is because areas outside the FOV still give a signal because the RF pulse(s) (and phase-encoding gradient) are applied to the whole head. However the value of the phase shift caused by the phase encode gradient will be outside the range of values assigned to cover the FOV, with the value assigned to a point just outside the FOV at one side of the image being identical to that assigned to a point just within the FOV at the other side. These points are therefore indistinguishable, and signal from objects at these points will overlap in the reconstructed image. This will cause a problem for analysis of atrophy if the wrap-around is large enough to overlay the brain (Figure 1-5f). In addition to wrap-around, if the FOV is very tight around the object being scanned then this can cause signal drop-out towards the edges of the image which may cause problems for segmentation of the brain or atrophy measurement techniques that rely on a constant intensity throughout the image. Both of these problems can be remedied by using a larger FOV.

Gradient non-linearity

Linear variation in the gradient field is required for accurate spatial encoding. However, gradient non-linearity may occur towards the edge of the imaging volume. This non-linearity can lead to distortions in the signal and geometry of images, which may affect the shape and boundaries of the brain, and as such influence the accuracy of cross-sectional volumes and longitudinal atrophy measures. This can be particularly apparent when the same subject is positioned differently between time-points, as the distortion is position dependent. As such, the distortion may be different between scans which can lead to error in measurements. Whilst scanner manufacturers supply software for the correction of linear variation, there is a discrepancy between the actual magnetic field and the gradient corrected for by the scanner software when non-linear variation is present. In

addition, most software works only in two dimensions which is not a complete solution. Recently three-dimensional algorithms have been applied to brain MRI and shown a significant improvement in distortion correction over two-dimensional methods (Jovicich *et al*., 2006).

1.6 Brain atrophy measurement methods

Although visual assessment of brain atrophy does not require specialist hardware or software, it is subject to reproducibility problems and is unlikely to be sensitive to small amounts of change. Quantitative approaches include manual methods that may be used for simple linear and area measurements, and automated or semi-automated software which is generally used for volume measurements to minimise operator-input time and increase reproducibility. This section provides an overview of methods that have been applied to MS subjects.

1.6.1 Cross-sectional or longitudinal?

Single time-point measurements of the length or width, area, and volume of whole brain and sub-regions have been utilised in studies of MS patients. Comparison with measurements from control subjects may indicate atrophy in these patients. However measurements based on a single scan can be difficult to interpret because of wide normal variability; normalisation to intracranial volume should be performed as small volume changes tend to be masked by the biological inter-individual variability in absolute brain size and volume. In addition, the progression and rate of atrophy can be assessed only indirectly from cross-sectional measurements, and one must assume that atrophy progresses linearly and at a similar rate between individuals. Gender and age, amongst other factors, have been shown to influence brain volumes and must be taken into account in analyses of cross-sectional data (Chard *et al*., 2002c). Longitudinal measurements allow disease progression to be monitored more precisely and the extent of true inter-individual differences identified. These may be obtained following the subtraction of serial cross-sectional measures (Equation 1.1), or by methods which directly measure the difference between ROIs following registration of images (Chapter 1.6.4).

$$Atrophy\ rate = \frac{V_1 - V_0}{t_1 - t_0} \tag{1.1}$$

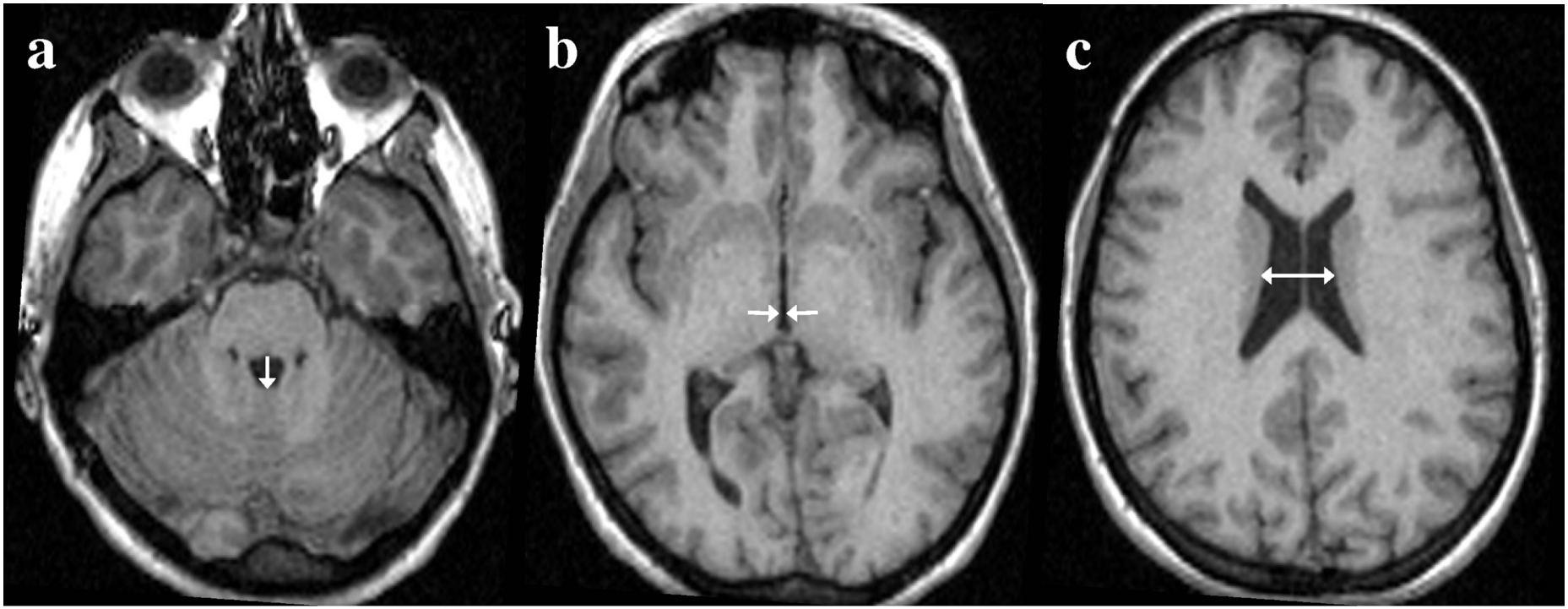
V_1 = volume on follow-up image, V_0 = volume on baseline image, $t_1 - t_2$ = interval

1.6.2 Two-dimensional brain measurements

Linear measures are the simplest atrophy measurements as they are quick to perform and can be applied to most structural MR acquisitions. Measurement of brain width on axially formatted MRI, has demonstrated atrophy rates of -0.64% year^{-1} in RRMS subjects with an intra-rater measurement coefficient of variation (CV, the standard deviation of measurements divided by the mean (σ/μ)) of 1% (Simon *et al*., 1999). Linear measurements of ventricular spaces, including the fourth, third and lateral ventricles, have been applied more widely and provide indirect assessment of brain atrophy (Simon *et al*., 1999; Turner *et al*., 2001) (Figure 1-6). Ventricular enlargement occurs as a result of tissue loss, and small losses can result in relatively large increases in CSF compartments. In cross-sectional studies, these measures may be normalised to brain size, by determining the brain width at the same level as ventricular width (Caon *et al*., 2003). One study measuring third and lateral ventricle width on axial MRI demonstrated annual increases of 4.5% and 5.5% respectively in MS subjects (Simon *et al*., 1999), and third ventricle width has been shown to correlate with third ventricle volume (Turner *et al*., 2001). Intra-rater CVs of 7% and 4% for third and lateral ventricle width measurements respectively have been shown (Simon *et al*., 1999).

Mid-sagittal measurements of corpus callosum (CC) area have also been performed in a number of MS studies. The CC is comprised of axonal tracts connecting the left and right brain hemispheres and has long been recognised as being particularly affected in MS. Cross-sectional analysis has shown significantly smaller average CC areas, by approximately 20%, in MS subjects relative to controls (Barkhof *et al*., 1998; Liu *et al*., 1999; Paolillo *et al*., 2000). Rate of atrophy in one of these studies was estimated to be -5.3% year^{-1} in patients with RRMS (Liu *et al*., 1999). Significant decreases in CC area have also been observed longitudinally (Martola *et al*., 2007; Pelletier *et al*., 2001; Simon *et al*., 1999), -4.9% year^{-1} in one study (Simon *et al*., 1999) and -1.8% year^{-1} in another (Martola *et al*., 2007). Methodological variability between studies in terms of measurement position, and the dependence of these two-dimensional measures on slice positioning and selection which are often based on subjective criteria (Benedict *et al*., 2004; Bermel *et al*., 2002; Sharma *et al*., 2004) limits these techniques however, and an intra-rater CV of 3% has been shown. Furthermore, slice selection may be harder when the head position and orientation within the MR scanner varies between patients and over time.

Figure 1-6 *Linear measures of a) the fourth ventricle, b) the third ventricle and c) the lateral ventricles.*



1.6.3 Cross-sectional methods

Delineation (segmentation) or classification of brain and CSF voxels on MRI allows global, regional or GM/WM volume quantification. Manual outlining of the brain is time-consuming, subjective and less reproducible than algorithms performing semi- or fully-automated segmentation. Numerous algorithms have been developed and may be based on thresholding (driven by the difference in brain/CSF or GM/WM signal intensity), region growing, clustering, deformable models or combinations of these. Some of the more commonly used methods and specific software packages that have been used in MS studies are discussed in this section.

Anatomatic

This semi-automated segmentation software uses a number of processes to obtain a volumetric estimate of the brain and GM/WM (Heinonen *et al*., 1997). Thresholds obtained from histogram analysis or defined manually are applied to images in order to classify skull, GM and WM. Region growing is then applied to the images to fill the segmentations and determine accurate intersections between the scalp, skull and CSF. The filled GM and WM images are subsequently combined with the skull image on a pixel-by-pixel basis with the aid of decision trees, which specify the voxels that should be retained from each image according to a set of rules (Heinonen *et al*., 1998). In this way classification of different tissue types within the image is achieved, including GM and WM, allowing investigation into the specific contributions of their pathology in MS. Misclassified brain lesions can also be identified on GM images prior to region filling and integrated into the final image according to rules designed for this purpose (Heinonen *et al*., 1998). This method has been successfully applied to T1 and T2/PD-weighted images (Ukkonen *et al*., 2003; Wu *et al*., 2007b), and based on known volume phantoms accuracy tests showed the error to be 1.5% of the total true volume (Heinonen *et al*., 1997). Although this method is semi-automated, manual implementation of thresholds and region growing applied on a slice by slice basis is recommended in some regions that are more difficult to segment. This method may therefore require considerable user input. Volumes must also be normalised to head size for cross-sectional analyses, but this is straightforward given that CSF is classified with the technique.

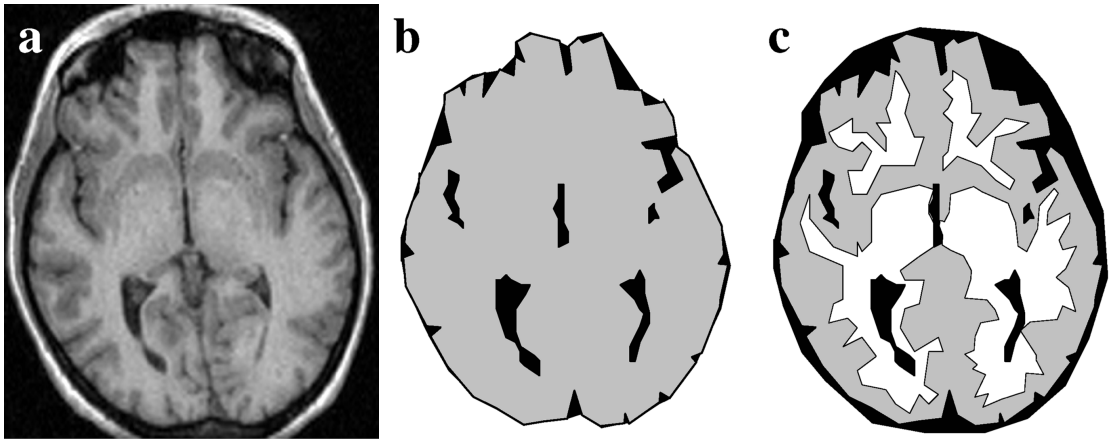
Brain parenchymal fraction

BPF is the ratio of brain parenchymal tissue volume to the total volume within the surface contour of the whole brain (Equation 1.2).

$$BPF = \frac{\text{brain tissue volume}}{\text{total volume within surface contour (brain tissue volume} + \text{CSF)}} \quad (1.2)$$

Whilst this term can be used to describe any technique that normalises brain volume measures in this way, a method designed to automatically determine brain and intracranial cavity volume was described by the Cleveland group and has been applied in numerous studies (Autoseg MS, Cleveland Clinic Foundation, Cleveland, OH) (Fisher *et al*., 1997; Rudick *et al*., 1999). This method includes CSF within the sulci but unlike many other methods, not that external to the outer brain surface (Figure 1-7). The method firstly applies an algorithm to a dual echo subtraction image to provide an initial brain segmentation based on optimal thresholding and connected components analysis. A 3D radial search operation is performed to detect the outer brain surface, non-brain structures are removed, and a smoothly contoured surface surrounding brain and CSF is produced. Within this surface contour, brain parenchymal tissue is separated from CSF by applying an optimal threshold. BPF is calculated using Equation 1.2.

Figure 1-7 *Segmentation for the brain parenchymal fraction and the brain to intracranial capacity ratio. a) original MRI, b) segmentation for calculation of brain parenchymal fraction, c) segmentation for calculation of brain to intracranial capacity ratio.*



Using the BPF, changes in voxel size, scanner gradient strengths or positioning within the scanner are minimised, as one segmentation is performed to obtain both brain tissue volume and normalising volume. This cancels out differences and improves reproducibility. Partial volume effects and intensity inhomogeneity are also accounted for during volume calculation, increasing accuracy. Scan-rescan reproducibility is the most stringent test of techniques. This fully automated method has a mean scan-rescan CV of 0.19% (Rudick *et al*., 1999) whilst mean absolute error of volume measurements performed on phantom images was less than 1.1% (Rudick *et al*., 1999). A semi-automated technique based on this original method has been applied to T1-weighted images (Bermel *et al*., 2003b; Sharma *et al*., 2004). The need for operator input to the algorithms and manual correction of the region makes these methods potentially more accurate but also more labour intensive and introduces greater variability. Relative to a phantom, accuracy has been estimated at 99.08%, but mean scan-rescan CV was higher than that of the original Cleveland method, ranging from 0.35% to 1.1% depending on image slice thickness (Bermel *et al*., 2003b; Sharma *et al*., 2004; Zivadinov *et al*., 2003; Zivadinov *et al*., 2004a). In addition mean intra- and inter-rater CV have been shown to vary between 0.03-0.37% and 0.31-1.00% respectively (Sharma *et al*., 2004; Zivadinov *et al*., 2003; Zivadinov *et al*., 2004a). One study directly comparing a fully- and semi-automated BPF method on different acquisitions found that the fully-automated technique did not provide satisfactory brain segmentation on the subtracted dual echo image in severely atrophied subjects, and that scan-rescan CV was greater on acquisitions with thicker slices (Horsfield *et al*., 2003). However the automated method used in this study was not the Cleveland algorithm.

Potential drawbacks of the BPF include possible insensitivity to increases in CSF spaces occurring with greater atrophy, peripheral atrophy being missed, and brain surface contour volumes changing over time thereby altering normalisation.

Brain to intracranial capacity ratio

Brain to intracranial capacity ratio (BICCR) is a measure similar to BPF, but includes extra cerebral CSF (between outer brain surface and dura) in addition to sulcal CSF (Figure 1-7). It has been used to detect atrophy from dual-echo T2/PD-weighted images (Brass *et al*., 2004; Collins *et al*., 2001). Each image is first registered into standard space (Talairach), to normalise for individual head size variations, before intensity

normalisation. A filter is applied to reduce noise within the image and improve voxel classification by a Bayesian classifier identifying GM, WM, CSF, lesion and background voxels. Mathematical morphology and masking is used to remove extracranial tissues and the total volume of voxels in each class is calculated and entered into Equation 1.3.

$$BICCR = \frac{\text{GM volume} + \text{WM volume} + \text{lesion volume}}{\text{GM volume} + \text{WM volume} + \text{lesion volume} + \text{CSF}} \quad (1.3)$$

Mean scan-rescan CV is 0.21% (Collins *et al*., 2001). WMF and GM fraction (GMF) may also be calculated. Like BPF, some scanner-related longitudinal variations, e.g. scanner gradient strength, are cancelled out. However the BICCR may be susceptible to partial volume effects, which could cause underestimation of CSF volume. This method also requires "training data" for the Bayesian classification of voxels which includes manual selection of approximately 50 voxels belonging to each tissue class in 20 subjects (Collins *et al*., 2001). Using the same training set for large numbers of scans may lead to biased results that do not take into account anatomical and physiological variability between subjects.

K-means clustering algorithm

Segmentation algorithms based on clustering utilise the data available in an image to iteratively characterise tissue properties and segment the image. One such automated method, based on the k-means algorithm (Goldszal *et al*., 1998), performs an initial segmentation of a skull-stripped image into voxel groups by minimising total intercluster variance with maximum likelihood estimation. Based on the mean intensity values of the tissue clusters an adaptive and iterative algorithm subsequently models and smoothes the estimated regions to obtain accurate segmentations of GM, WM and CSF by classifying each pixel into the class with the closest mean. In tests of accuracy on a phantom the method yielded errors (as a percentage of the total volume) of 0.43% for brain volume, 1.96% for GM volume and 2.71% for WM volumes. Mean scan-rescan absolute difference based on three subjects was 0.31% for brain volume, 0.71% for GM volume and 0.75% for WM volume (Goldszal *et al*., 1998). In addition to providing estimates of tissue classes, another advantage of this approach is that image inhomogeneity can be simultaneously corrected for. Whilst this technique is able to quantify separate GMF and WMF, additional processing must be performed in order to correct these volumes for

lesion misclassification. This method has successfully been applied to T1 and dual echo subtraction images, however results from FLAIR images were unsatisfactory (Leigh *et al*., 2002). Disadvantages of this technique include the need for a separate skull stripping procedure, which in the original method (Goldszal *et al*., 1998) removed an undetermined amount of sulcal CSF, making it difficult to accurately normalise brain volumes. In addition it is thought that the segmentation algorithm may not perform adequately on the cerebellum (Goldszal *et al*., 1998).

Fuzzy connected principles (FCP)

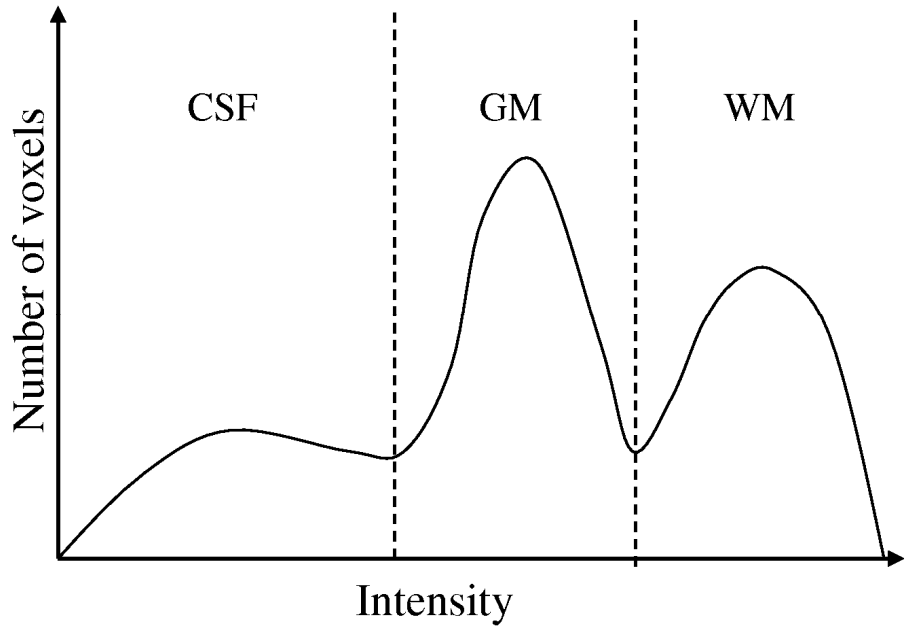
Another algorithm based on clustering and using the theory of "fuzzy connectedness" (Pham & Prince, 1999; Udupa & Samarasekera, 1996) has been applied to segment brains on different acquisitions and quantify atrophy in MS (Ge *et al*., 2000b; Leigh *et al*., 2002). As with the k-means algorithm initialisation is required, which in this case is an operator identifying points of GM, WM and CSF within dual echo images, each of which is then automatically detected as a 3D fuzzy-connected object. Subtraction of the dual echo images is effective in obtaining a CSF-only image. Voxels can belong to multiple classes with varying degrees of membership, allowing greater information to be retained from the original image, and dealing with partial volume effects. Again, developments to the algorithm have allowed intensity inhomogeneity to be corrected for simultaneously, albeit at the cost of greater computational time (Pham & Prince, 1999) and lesions can also be identified semi-automatically during the procedure. Scan-rescan CV for whole brain measured on dual-echo images was 0.23% (Ge *et al*., 2000b), whilst intra-rater and inter-rater reproducibilities of 0.38% and 0.68% respectively have been shown (Leigh *et al*., 2002). Scan-rescan CV for GMF and WMF respectively are 2.1% and 1.9% (Ge *et al*., 2001). Although this method has been applied to T1-weighted images with similar results, application to FLAIR images resulted in greater inter-rater reproducibility (Leigh *et al*., 2002). The simultaneous generation of CSF volumes allows for calculation of normalised volumes.

Histogram segmentation

Histogram segmentation algorithms are based on the intensity distribution of voxels within an image (Figure 1-8). One scheme, optimised for T1-weighted images (Kovacevic *et al*., 2002; Leigh *et al*., 2002), initially requires a dual echo image histogram to determine the optimal thresholds for separation of brain from non-brain voxels. An

automated spatial connectivity algorithm is applied to refine the classification and manual editing can be performed where necessary. The extracted brain region is used to mask the corresponding T1-weighted image following registration, and classify brain voxels as GM, WM and CSF, based on the Gaussian distribution of voxel intensities. Partial volume voxels are assigned according to a weighting factor. Separate skull-stripping must be performed prior to image analysis. A scan-rescan error of 0.13% of total intracranial capacity has been reported whilst mean absolute scan-rescan differences in proportional tissue volume were 0.8% and 1.3% for GM and WM respectively (Kovacevic *et al*., 2002). One of the disadvantages of this particular method is the requirement of both T1-weighted and dual echo images.

Figure 1-8 *Example of an intensity histogram.*



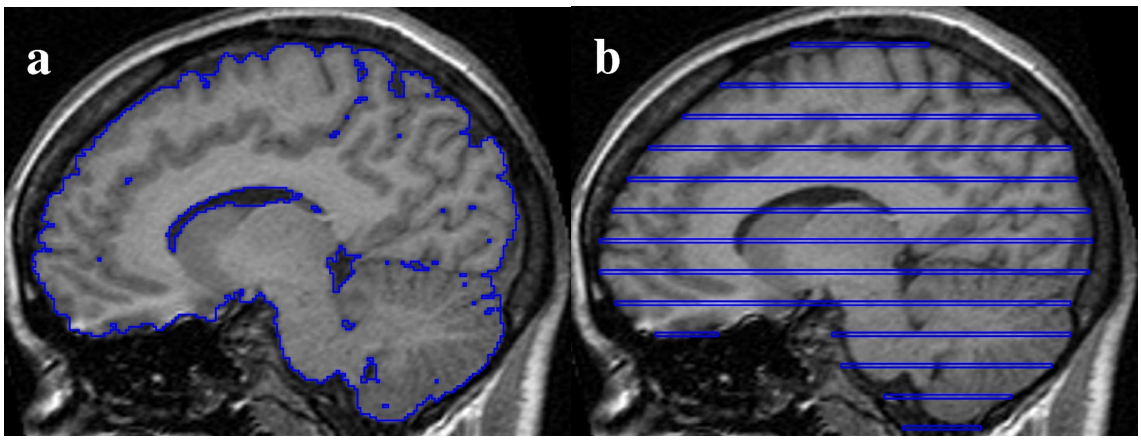
Similarly MeVisLab Brain Volumetry software uses automated regional histogram analysis to derive whole brain, WM, GM and intracranial volume from 3D images following skull stripping (Lukas *et al*., 2004). Image noise and partial volume effects are taken into account which will increase the accuracy of this technique. Based on six subjects scan-rescan CV was 0.3% for whole brain volume (WBV), 1.1% for GM and 1.7% for WM. However suboptimal repositioning of subjects resulted in image inhomogeneity and considerable increases in CV for GM and WM (2.1% and 3.4% respectively), demonstrating the reliance of this method on consistent placement of the patient in the scanner or pre-processing methods to remove intensity inhomogeneities.

Based on phantoms, absolute error for whole brain, GM or WM was shown to be less than 5ml for each volume.

MIDAS

This interactive software (Medical Image Display and Analysis System) enables semi-automated 3D analysis of T1-weighted MR images based on intensity thresholding, region growing and morphological operators (Figure 1-9) (Freeborough *et al*., 1997). To obtain an initial approximation of the brain an operator must select two thresholds that represent the range of brain voxel signal intensities, in addition to the most inferior point in the brain. Conditional erosion(s) and dilation(s) of the resulting region are performed, with thresholds defined to prevent erosion of WM or dilation into points outside the intensity of brain tissue. Rethresholding is subsequently performed in order to reclassify brain voxels that have been removed by the previous steps, for example thin structures like the fornix. Mean absolute error as a proportion of brain volume was estimated to be 0.34% (Freeborough *et al*., 1997). A CV of 0.46% for intra-rater and 0.54% for inter-rater reproducibility have been shown (Fox *et al*., 2000b). This technique simultaneously skull-strips the image and performs brain segmentation, however partial volume effects are not taken into consideration with this technique and normalisation for head size must be performed separately through estimation of total intracranial volume (Figure 1-9) (Whitwell *et al*., 2001).

Figure 1-9 *MIDAS segmentation of a) brain, b) total intracranial volume, based on intensity thresholding.*



Region growing

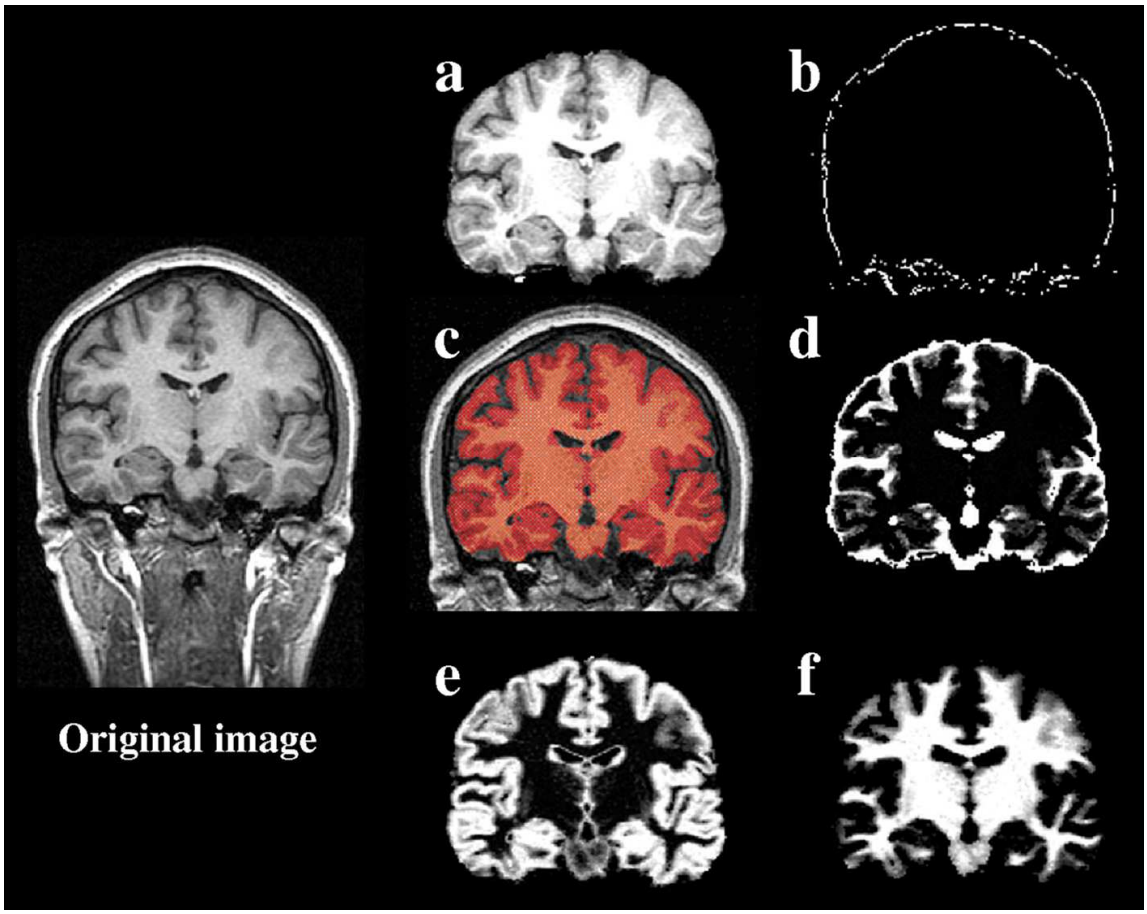
Semi-automated segmentation algorithms based on region growing require manual positioning of a seed in any part of the brain parenchyma, and an ROI is grown from this containing all connected pixels until an edge in the image is met (Gasperini *et al*., 2001; Kalkers *et al*., 2002; Rovaris *et al*., 2000). This condition is usually based on local thresholds and the lower threshold can be automatically determined by an edge detection filter, designed to perceive strong intensity gradients. However both upper and lower thresholds can be changed manually on a slice-by-slice basis and boundaries can be drawn to limit the ROI. A mean intra-observer CV of 1.9% was demonstrated on T1-weighted images (Rovaris *et al*., 2000). This technique may be subject to reproducibility problems with manual implementation of thresholds, requires separate normalisation for head-size in cross-sectional studies, and may take considerable time for 3D volumetric acquisitions if segmentation is performed on a slice-by-slice basis. Due to the complex structure of the brain and partial volume effects, this method can lead to distinct regions becoming connected, or extracted regions containing holes or becoming disconnected.

SIENAX

SIENAX (Structural Image Evaluation, using Normalisation, of Atrophy – Cross-sectional) (Smith *et al*., 2002) uses a fully automated algorithm, based on a Markov Random Field model, to estimate volume measurements of whole brain, GM and WM (Figure 1-10). An automated algorithm, the Brain Extraction Tool (BET) (Smith, 2002), is used to extract brain from non-brain and estimate the outer skull surface. This is based on histogram analysis to find an approximate brain/non-brain threshold followed by a deformable model-based technique using triangular tessellation of the surface of a sphere. The brain image is registered to standard space (based on the Montreal Neurological Institute (MNI) standard template, MNI-152) using the estimated skull surface to constrain scaling, thereby normalising for head size, and a brain mask is applied to exclude extracerebral tissue. SIENAX can segment the extracted brain into GM, WM and CSF and includes partial volume modelling and intensity inhomogeneity correction thereby increasing measurement accuracy (Zhang *et al*., 2001a). However for MS subjects, separate lesion classification must be performed and included as a mask in the process. Mean scan-rescan whole brain volume error is 1%, which has been shown to be independent of slice thickness (Smith *et al*., 2002). However a study found this method to be less accurate than semi-automated methods of brain extraction (Hahn *et al*., 2004) and

adjustment of parameters may be required, particularly placement of the initial deformable model. The procedure can be applied to T1- or T2-weighted images.

Figure 1-10 *SIENAX segmentation. a) BET extracted brain, b) BET extracted estimation of skull, c) estimation of brain volume, d) estimation of CSF volume, e) estimation of grey matter volume, f) estimation of white matter volume.*

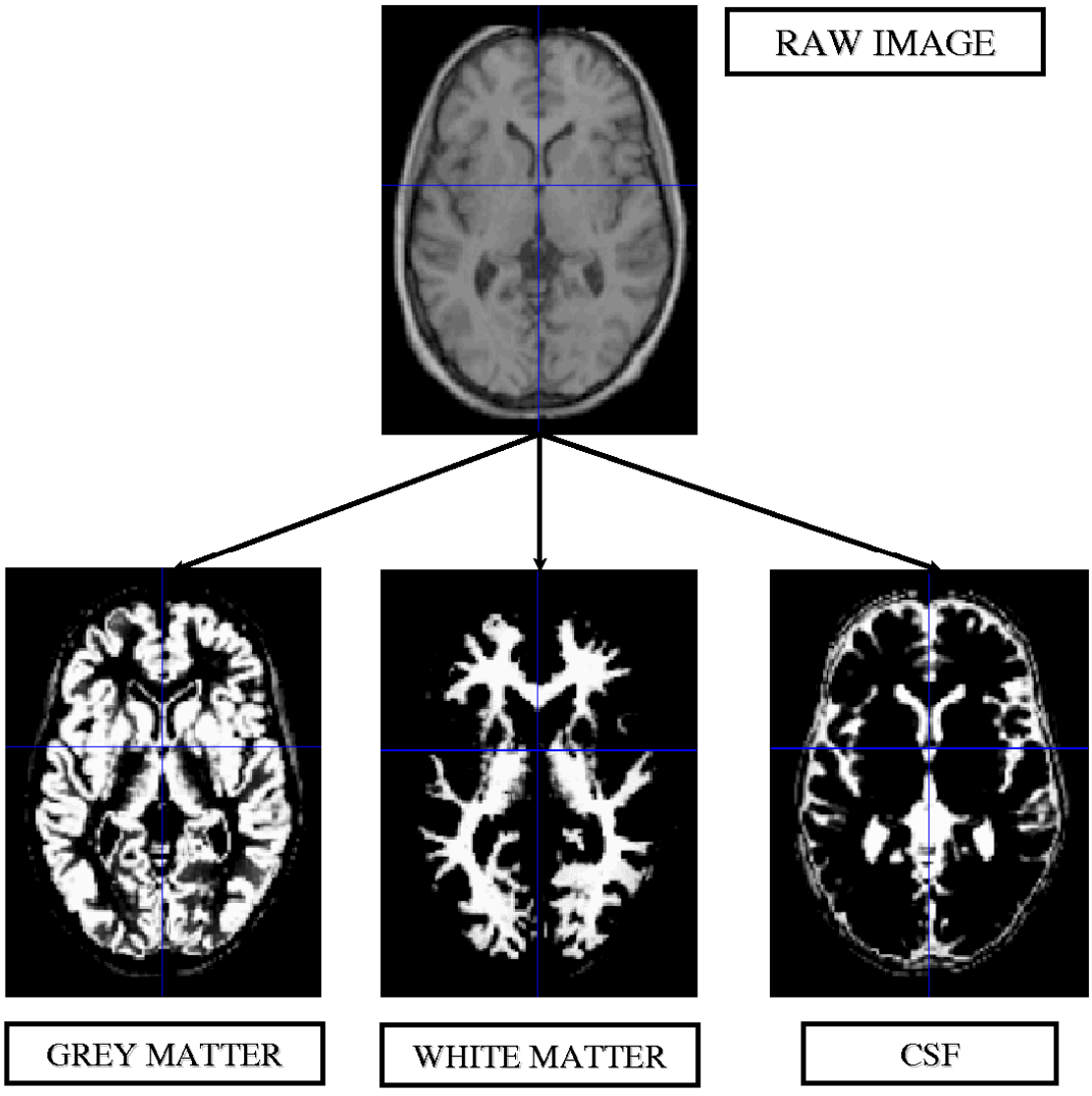


SPM brain parenchymal fraction

SPM software (Statistical Parametric Mapping; Wellcome Department of Cognitive Neurology, Queen Square, London) (Ashburner & Friston, 1997; Ashburner & Friston, 2000) is a widely used package originally applied to functional imaging studies. Within SPM, methods exist to classify MR image voxels into GM, WM and CSF. Before segmentation all images are placed into stereotactic space (based on the MNI-152 standard template), corrected for intensity inhomogeneity and masked to remove extracranial tissue. Segmentation is based on a stereotactically normalised *a priori* atlas (from a database of normal brain images) and image intensity thresholds, and mutually exclusive masks are generated for each tissue class (Figure 1-11). Misclassification of

WM lesions as GM or CSF, may introduce error to quantification, and the problem has been approached in two different ways. One approach is to delineate lesions manually and apply this information to override all SPM tissue classifications. A BPF-like measure is calculated from the sum of GM, WM and lesion volumes divided by the sum of GM, WM, lesion and CSF volumes (Chard *et al*., 2002b). Alternatively, each mutually exclusive mask is subjected to a morphological erosion followed by a conditional dilation where only voxels previously classified as GM or WM are dilated. Resulting GMFs and WMFs are regarded as corrected for lesions, and BPF is calculated from the sum of GM and WM volumes divided by the sum of GM, WM and CSF volumes (Kassubek *et al*., 2003). However it is unclear how accurate this approach may be.

Figure 1-11 *SPM segmentation*.

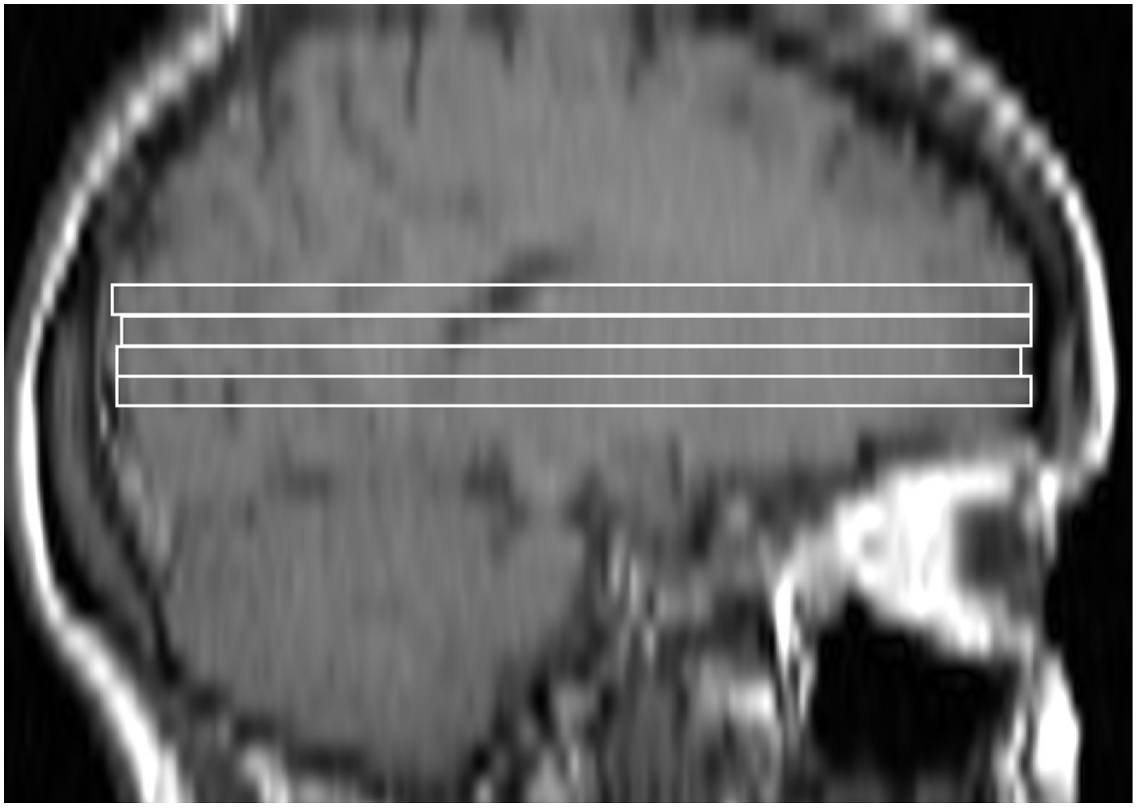


Although SPM has been applied to acquisitions with different slice thicknesses, it has been suggested that acquisitions with thick slices (5mm) may result in poorer segmentations (Sharma *et al*., 2004). Scan-rescan CVs of 0.5% for BPF, 0.7% for GMF and 1.1% for WMF have been shown on 3D T1-weighted control images (Chard *et al*., 2002c) whilst mean intra- and inter-rater CVs of 0.09% and 0.19% respectively have been reported (Zivadinov *et al*., 2004a).

Central cerebral volume (CCV)

This method attempts to quantify brain atrophy on acquisitions with thick slices, by measuring the volume of four to seven (dependent on image slice thickness) contiguous axial slices from the central portion of the brain (Figure 1-12). The most caudal slice is chosen at the level of the velum interpositum cerebri, which is thought to be a stable landmark in the presence of atrophy, thus reducing measurement error that could present over longitudinal analyses (Gasperini *et al*., 2002; Ingle *et al*., 2002; Losseff *et al*., 1996).

Figure 1-12 *Central cerebral volume measured over four axial slices (5mm thick).*



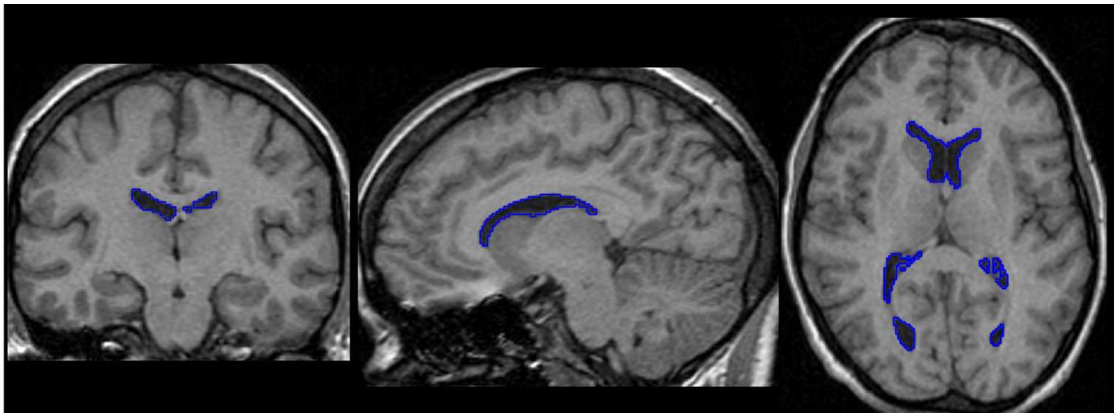
The original automated method was applied to selected slices in order to extract the brain from skull and CSF but other algorithms could be used to determine this measure. The

algorithm involves discriminant analysis of the intensity histogram to identify the optimal threshold separating brain from non-brain, creating a binary image of background/brain. Erosion of the binary image, followed by dilation separates brain from other extracerebral tissue. The binary image is used to mask the original image, producing the segmented brain, which can be manually edited if required. A four-slice (5mm) measure scan-rescan CV was 0.56% (Losseff *et al*., 1996). Although only a limited number of slices are used to estimate cerebral atrophy, MS lesions are often located in the selected region which also includes a large proportion of the lateral ventricles and cortical sulci where atrophy is often qualitatively prominent. In addition segmentation may be less time consuming than other semi-automated methods segmenting the whole brain. However, although this technique may show larger percentage losses of tissue than whole brain measures it may be less reproducible due to differences in acquisition (slice thickness), and subject repositioning and orientation.

CSF measures

The high contrast boundary between CSF and brain tissue on some MR acquisitions allows highly accurate identification of ventricular borders using intensity-based techniques. Semi-automated thresholding and region growing techniques, such as MIDAS and Anatomatic, have been used to outline the lateral ventricles and temporal horn regions (Dalton *et al*., 2002a; Dastidar *et al*., 1999; Fox *et al*., 2000b; Kalkers *et al*., 2002) (Figure 1-13).

Figure 1-13 *MIDAS segmentation of the lateral ventricles.*



Reported intra-rater CV for the MIDAS technique ranges from 0.02% to 0.89% (Brex *et al*., 2000; Dalton *et al*., 2002a; Fox *et al*., 2000b), whilst an inter-rater CV of 0.32% has

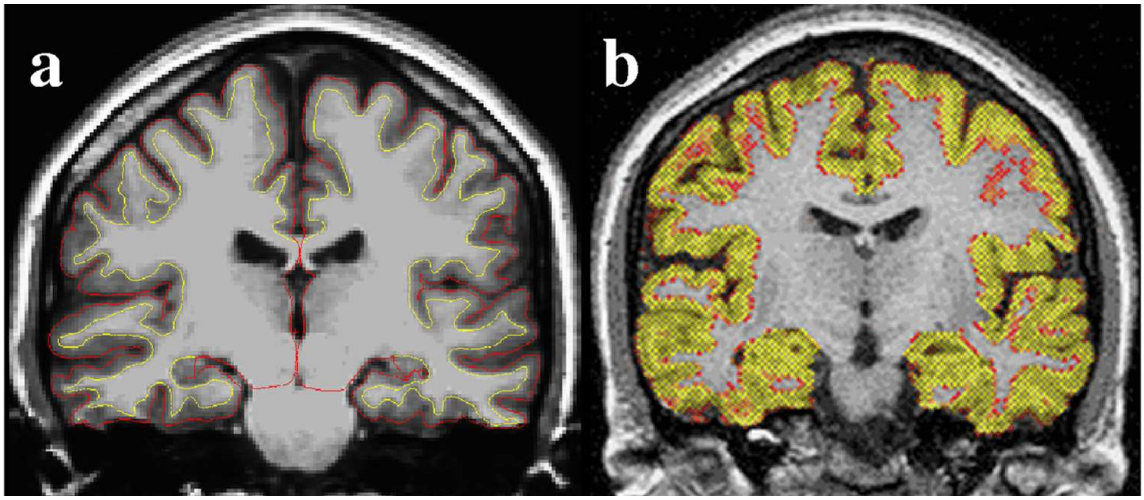
been reported (Fox *et al*., 2000b). SIENAX is also able to automatically obtain an estimate of ventricular CSF volume. Whilst SIENAX is fully automated, semi-automated thresholding techniques are quick and easy to perform given the clear high-contrast boundary between brain and CSF.

Cortical grey matter

Recently, automated methods for measuring cortical thickness of the entire brain and estimating cortical GM atrophy have been developed. Knowledge of the regional distribution and evolution of GM atrophy may provide useful information on the pathogenesis of MS. Two methods that have been applied in MS, Freesurfer and CLASP (Constrained Laplacian Anatomic Segmentation using Proximity), reconstruct the cortical surface from volumetric MRI (Dale *et al*., 1999; Kim *et al*., 2005). Deformable surface algorithms are used to obtain estimates of the GM/WM and GM/CSF interfaces which are accurate at the subvoxel level (Figure 1-14a). The thickness of the cortex is computed at each point within the surface, and global and regional mean thicknesses can be determined with a high level of sensitivity and accuracy. Based on scan-rescan data from one control subject, points on the cortical surface were matched and the mean standard deviation of the measures from each of these points was 0.25mm in one method (Fischl & Dale, 2000). In addition, this study found that over 99% of measures across the cortical surface were within the known bounds of 1-4.5mm. Individual mean cortical thickness can be computed, in addition to statistical analyses of group differences (Sailer *et al*., 2003), which may aid identification of regional cortical atrophy. Disadvantages of this technique include the fact that good GM/WM contrast is necessary for these analyses and that the GM/WM surface is deformed to create the GM/CSF surface, therefore WM segmentation errors may be propagated. Although errors can be corrected manually, including lesions that have been identified as cortex, this may be time-consuming and will decrease the reproducibility of the technique.

Based on the GM segmentation obtained, SIENAX will also quantify cortical GM volume and has the advantage that it is fully automated (Figure 1-14b) (Smith *et al*., 2002). Measures will be subject to the same advantages and disadvantages as other measures obtained using this software.

Figure 1-14 *Freesurfer and SIENAX cortical segmentations. a) Freesurfer estimates the grey matter/white matter surface (yellow line) and the grey matter/CSF surface (red line) to determine cortical thickness, b) SIENAX estimates cortical grey matter volume (yellow region).*



Regional analysis

From visual inspection of MR images, atrophy appears to be a global phenomenon in MS. However regional analysis of the caudate, cerebellum, cerebral hemispheres, frontal and temporal lobes, and thalamus have been investigated (Benedict *et al*., 2005; Bermel *et al*., 2003a; Cifelli *et al*., 2002; Filippi *et al*., 1998; Liu *et al*., 1999; Zivadinov *et al*., 2003). Segmentation of these regions has usually involved manual outlining or volume estimates using point counting based on the Cavalieri method (Liu *et al*., 1999). A semi-automated parcellation method was reported in 2004 (SABRE – Semi-Automatic Brain Region Extraction) which divides each brain hemisphere into 13 regions taking into account anatomical divisions (separation of cortical lobes), identified areas of interest, and optimisation of reliability and efficiency (Dade *et al*., 2004). Images must be pre-processed to remove extracerebral tissue and align scans before 15 landmarks are manually identified. These provide co-ordinates for an individual Talairach atlas grid to be transformed to an image, from which the algorithm automatically delineates the 26 brain regions. This method has been used in conjunction with tissue compartment segmentation software (Kovacevic *et al*., 2002) to provide regional volumes of GM and WM. Whilst intra-rater correlation coefficients ranged between 0.95 and 0.99 depending on the region, it is unclear how accurate this method is. However a validation study did

demonstrate significant overall loss in regional GM and WM volumes in old compared with young normal control subjects (Dade *et al*., 2004).

SPM software has been utilised to perform voxel based morphometry (VBM), a fully automated whole-brain technique, allowing unbiased analysis of regional differences in tissue density between subject groups from structural imaging (Ashburner & Friston, 2000). Images are registered to a template and robust statistical techniques are used to make voxel-wise group comparisons over the whole brain without the need for regional segmentations relying on *a priori* assumptions. It can provide insights into characteristic patterns of atrophy and anatomical differences between subject groups. However VBM may be less sensitive to changes between groups in areas with high natural variance.

1.6.4 Longitudinal registration-based methods

Detection of small diffuse brain volume changes from serial MRI is difficult using methods that rely on outlining of the brain, because the results are critically dependent on the reproducibility of segmentation. Image subtraction is an alternative method of assessing diffuse atrophy from serial scans. Direct quantification of volume change is subject to less error than quantifying and subtracting brain volumes at different time-points where errors may occur in measurements at both time-points. For image subtraction to produce meaningful results however, serial images must be positionally registered (spatially matched) (Figure 1-15).

Brain boundary shift integral

The brain boundary shift integral (BBSI) (Fox & Freeborough, 1997) has been used on serial 3D T1-weighted images to calculate atrophy from difference images (Fox *et al*., 2000b). Semi-automated segmentation of baseline and repeat brain regions allows brain registration and provides an estimate of the brain boundary region. An automated registration algorithm determines the rotations, translations, scalings and shear that are required to obtain a subvoxel match over the whole brain, and a linear scaling is used to account for variations in voxel size due to scanner drift. Atrophy quantification is based on integrating the sampled difference in brain voxel intensities between the baseline and registered repeat image and represents the total volume traversed by the brain/CSF boundaries in going from baseline to registered repeat scan (Figure 1-16).

Figure 1-15 *Registration of serial brain MRI for atrophy quantification. a) Registration of serial images allows volume change to be quantified directly by looking at the difference image, b) A difference image (coronal plane) of an MS patient following registration of serial brain MRI (one year interval), showing change particularly around the ventricles (arrow).*

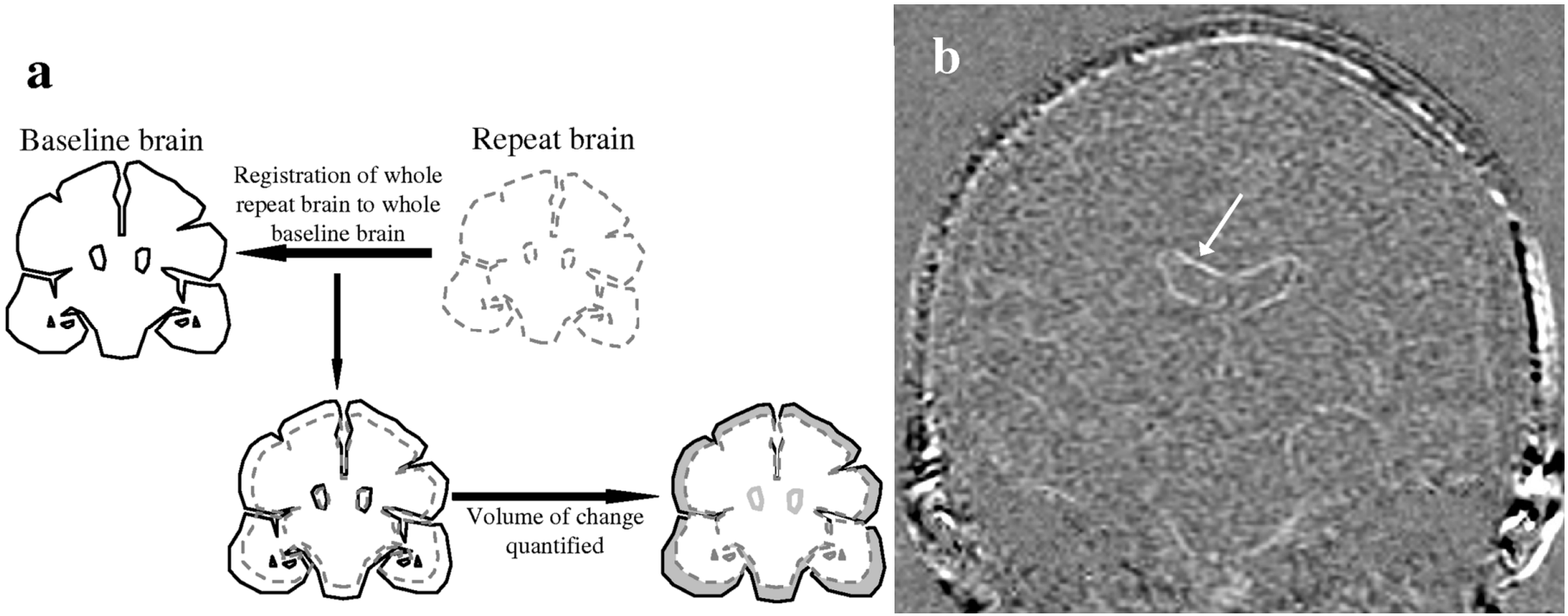
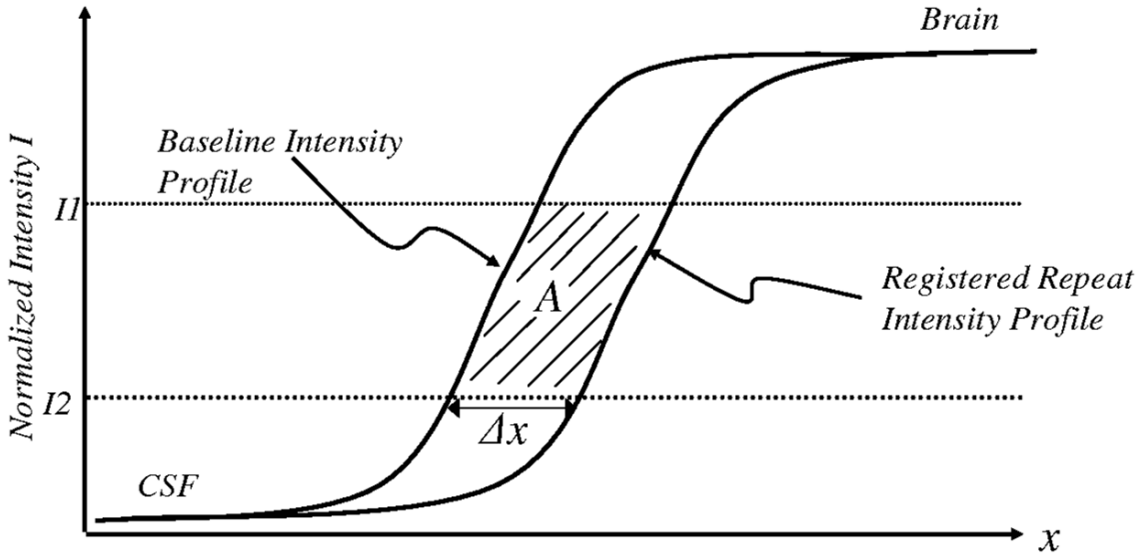


Figure 1-16 *Calculation of the brain boundary shift integral. The figure shows a one-dimensional representation of the intensity profile through a brain boundary on serial imaging. The boundary shift (Δx) is approximated as the area A divided by ($I_1 - I_2$).*



Slight segmentation errors, small positional shifts, or shape changes should not affect the BBSI. Mean absolute error on scan-rescan testing was estimated to be approximately 0.13% of mean brain volume and comparison of the BBSI with simulated volume loss yielded correlation coefficients of 1.000 (Fox & Freeborough, 1997). The BSI has also been applied to quantify ventricular enlargement directly (Freeborough & Fox, 1997). As the BSI relies on an intensity transition between different tissues, the contrast of images must be consistent across serial imaging. Changes in voxel intensity between a baseline and repeat image could be incorrectly interpreted as atrophy or "growth" of the brain. Differential bias correction (DBC), has been described for the correction of differences in the bias field between two images to improve the precision of atrophy measurement (Lewis & Fox, 2004).

SIENA

Another automated registration-based atrophy measurement method SIENA (Structural Image Evaluation, using Normalisation, of Atrophy) (Smith *et al.*, 2002), is the longitudinal version of SIENAX. Similarly, it achieves segmentation of the brain using a deformable tessellated mesh to model the brain surface. Estimation of the outer skull surface, in the case of SIENA, is used to constrain the registration of serial images whilst normalising for imaging geometry changes. The brain surface is detected using a robust

method that generates a full tissue-type segmentation, and percentage brain volume change (PBVC) is based on the movement of this edge between images, with sub-voxel accuracy. Edge finding is relatively insensitive to changes of intensity in tissues through serial images, and this technique can be applied to both T1- and T2-weighted acquisitions, and images with different slice thicknesses (Smith *et al*., 2002). Median absolute scan-rescan error for brain volume change was reported to be 0.15% (Smith *et al*., 2002). Developments to this software have made it possible to perform voxelwise group statistical analysis, potentially enabling identification of areas that preferentially atrophy in MS. However this has not been employed in studies to date, possibly because changes at the edges of the brain may not be specific to atrophy in that region.

Non-linear registration methods

Affine rigid-body registration of serial images followed by a non-linear registration can also be used to assess atrophy directly, but studies using these methods in MS are lacking to date. The non-linear registration transforms the rigidly-registered repeat image to match the baseline and at each voxel a Jacobian matrix can be obtained that describes the deformation (Freeborough & Fox, 1998; Rueckert *et al*., 1999; Shen & Davatzikos, 2003). Each voxel is considered as expanded or contracted and the Jacobians can be summed over a previously derived region (e.g. brain) in order to obtain an estimate of atrophy.

A method that utilises non-linear registration in order to propagate segmentations from the baseline image to repeat images was developed by Calmon *et al.* (Calmon & Roberts, 2000). This technique firstly registers serial images using an automated rigid-body registration algorithm that detects crest-lines in the images and matches points corresponding to a maximum curvature in the principle directions. Intensity scaling is performed so that serial images are the same average intensity and an automated non-linear registration is used ("demons" method) to calculate the residual deformations between serial images that are not accounted for by rigid-body registration. The deformation field is applied to a segmentation of the baseline image which is automatically deformed and propagated through any number of serial images to provide an estimate of volume change. Reported scan-rescan CV was 0.5% for this technique, which has also been successfully applied to the lateral ventricles with a CV of 1% (Calmon & Roberts, 2000).

Regional group-based analysis

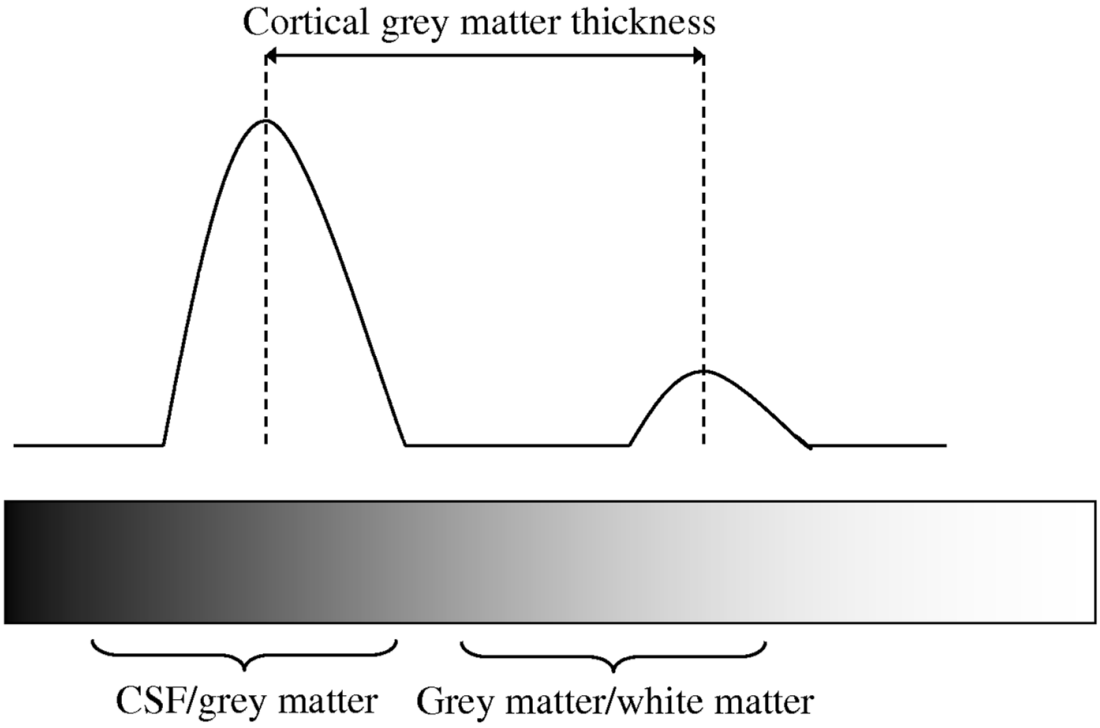
SPM and VBM can be utilised with other longitudinal image analysis techniques to investigate regional differences in atrophy rates (Ashburner & Friston, 2000). Non-linear registration can be used to localise changes within individuals and VBM can be used to determine consistent changes within groups, although this method has not been applied in longitudinal studies of brain atrophy in MS to date (Ashburner & Friston, 2000; Scahill *et al*., 2002). Images must be spatially normalised and smoothed prior to analysis. As with cross-sectional VBM, it may provide insights into characteristic patterns of atrophy between subject groups, without *a priori* assumptions, whilst in addition it is likely to be less influenced by the natural morphologic variability between subjects.

SPM statistics have also been utilised with SIENA in one study of MS (Pagani *et al*., 2005b). For each subject, scalar values from each boundary point calculated by SIENA were saved as displacement maps. Following spatial normalisation and smoothing of these maps group analysis was performed using SPM.

Cortical thickness

Another method of computing cortical thickness is based on analysis of the derivative of image intensity profiles which allows determination of GM/WM and GM/CSF interfaces at the subvoxel level (Chen *et al*., 2004) and is an extension of the SIENA technique. The derivative of the intensity profile is affected only by the rate of change of the intensity and cortical thickness is estimated as the difference in the maxima of this profile (Figure 1-17). This technique is not reliant on high-resolution MRI (images with 3mm thick slices have been successfully analysed) and does not require segmentation of images, thereby reducing the errors associated with this process. It also provides a measure of the integrity of the GM/WM interface. Although this method can be performed for the cross-sectional analysis of images, several features of the sampling method used to create intensity profiles make it more robust for longitudinal analysis. Firstly, sampling is performed at the crowns of the gyri as opposed to in the sulci. As gyri generally have thicker cortex this may mean there is bias towards greater measurements; indeed the study by Chen *et al.* showed a trend for greater thickness measures in patients with MS compared with post-mortem studies. Secondly, sampling is performed on a point-by-point basis over the exposed cortical surface and this will vary from subject to subject. Investigating change within a subject from registered imaging overcomes these potential problems.

Figure 1-17 *Calculation of cortical thickness from the derivative of the intensity profile perpendicular to the brain boundary. Cortical thickness is the distance between the two maxima on this profile. (Reproduced from Chen et al., 2004).*



Longitudinal methods within Freesurfer exist and are based on using processed results from cross-sectional data to analyse later time-points. This mainly involves initialisation of the processing of longitudinal data sets using the processed results from another time-point, following registration of images.

1.6.5 Effect of lesions on brain atrophy measures

It is possible that lesions may affect segmentation- and registration-based brain atrophy measurements, for example T1-hypointense lesions could be misclassified as CSF, or GM lesions might cause subtle signal intensity changes that affect segmentation. However analysis of ten MS patients with high T1 lesion loads found that lesion misclassification had a negligible effect on BPF measurements from SPM (Sharma *et al*., 2004). Likewise, another study showed no significant differences in tissue volumes calculated from SPM, between images with simulated WM lesions and those without, although GMF was slightly higher and WMF slightly lower when lesions were present (Chard *et al*., 2002c; Dalton *et al*., 2004). It may be concluded that brain volumes derived using SPM segmentations are relatively insensitive to WM lesions, but it is unclear how other

measurement techniques may be affected. Measurement of GM atrophy may provide a more direct assessment of neurodegeneration in MS, unhindered by fluctuations in tissue volume associated with inflammation.

2 Brain atrophy in multiple sclerosis

2.1 Clinically isolated syndromes

Sixty to eighty percent of patients presenting with CIS suggestive of MS (e.g. optic neuritis) develop clinically definite MS, and the proportion is greater in those who have MRI-visible brain lesions (Brex *et al*., 2002).

Studies have demonstrated significantly greater ventricular enlargement within one year in people who develop MS compared with those who remain stable (+0.3 to 0.8cm^3 year^{-1} compared with -0.1 to +0.06cm^3 year^{-1}) (Brex *et al*., 2000; Dalton *et al*., 2002a; Dalton *et al*., 2006). A three year follow-up of 58 CIS subjects also found that in 31 subjects developing MS, ventricular volume increased by a mean of 38.9%, whilst in 27 subjects remaining stable only a 5.4% increase was observed (Dalton *et al*., 2004).

Brain atrophy has been measured in CIS using SIENA. A study of 31 subjects in which the follow-up period was only four to six months found a -0.27% (standard error 0.16%) loss of brain volume (Filippi *et al*., 2003). Assuming a linear atrophy rate this was equated to a -0.69% year^{-1} brain volume loss, however because of the short interval there is a wide 95% confidence interval (CI) on the annualised rate of loss. Median estimated atrophy rate was only -0.3% year^{-1} (SD 0.6) in another study of 20 CIS subjects studied over one year (Rovaris *et al*., 2003), and SPM measures of BPF, GMF and WMF showed a small non-significant decrease over this period (Agosta *et al*., 2006). Another study, from the same author, of 35 CIS subjects, which may have included some of the same patients, found an annual brain atrophy rate of -0.41% year^{-1} (SD 0.58) (Rovaris *et al*., 2005b). None of these studies grouped patients according to clinical follow-up or inflammatory activity however, which may have increased variability in the quantifications.

Analysis of 38 CIS subjects who remained relapse free after an 18 month follow-up period, showed a median -1.1% (interquartile range (IQR) -1.91 to -0.67) loss of brain volume over this period (Paolillo *et al*., 2004). However when subdivided into patients with (n=25) and without (n=13) at least one new active lesion during the first six months of study, a significant difference was found between atrophy rates: -1.71% in active and

-0.48% in inactive patients over 18 months. Likewise in the placebo arm of a trial of interferon beta-1a in CIS, which included over 100 subjects, brain volume loss was -0.83% year^{-1} (SD 1.09) (Filippi *et al*., 2004). Importantly, atrophy rates were greater in subjects who developed clinically definite MS compared with those who did not. A cross-sectional study grouping CIS patients according to whether there was evidence of dissemination in space of lesions at presentation (10 patients without (CIS) and 32 with ("probable" MS), both groups with a mean disease duration 0.7 years), found that BPF, GMF and WMF were all reduced in the "probable" MS group compared with controls and CIS subjects (Calabrese *et al*., 2007a). Measures of cortical thickness were also lower in the "probable" MS group compared with the CIS group, 2.22mm (SD 0.09) versus 2.51mm (SD 0.11). Similarly a longitudinal analysis using SPM to analyse BPF, WMF and GMF grouped subjects presenting with a CIS into those who met the McDonald criteria at three year follow-up and those that did not (Dalton *et al*., 2004). Significant decreases in both MS (31 subjects) and CIS (27 subjects) groups were observed at three years: respectively -1.4% and -0.6% in BPF and -3.3% and -1.1% in GMF. The decreases were significantly greater in the MS group, which also showed a weak but significant 1.3% increase in WMF that was suggested possibly to have been due to inflammation.

2.2 Relapsing remitting multiple sclerosis

Many cross-sectional studies have shown that brain volume is reduced in RRMS subjects compared with age-matched controls (Bermel *et al*., 2003b; Collins *et al*., 2001; De Stefano *et al*., 2003; Kalkers *et al*., 2001a; Lin *et al*., 2003; Paolillo *et al*., 2000; Traboulsee *et al*., 2003). The majority of longitudinal studies estimate atrophy rates of around -0.7 to -1.5% year^{-1}; these rates of loss are seen even in those subjects at the earliest stages of disease prior to significant disability (Chard *et al*., 2004; Rovaris *et al*., 2000) (Table 2-1). Mean atrophy rate in 34 subjects, estimated using SIENA, found a -0.7% year^{-1} (SD 0.9) loss in brain volume over an interval of one year (Rovaris *et al*., 2003). In concordance, another study using SIENA on RRMS subjects with a similar mean disease duration found a volume change of -0.91% over an 18 month interval (Oreja-Guevara *et al*., 2005). Other measurement methods show similar findings; BPF volume loss was -0.7% year^{-1} (IQR -1.3 to -0.01) in 42 subjects with RRMS (Kalkers *et al*., 2002) and analysis of 3D fast spoiled gradient recall (FSPGR) images found that over two years BPF volume loss was -1.5% in 21 RRMS subjects compared with -0.6% in controls (Tiberio *et al*., 2005). In addition, application of the BBSI to 3D FSPGR images

showed a median atrophy rate of -0.8% year^{-1} (IQR -0.9 to 0.2) in RRMS subjects which was greater than the control rate (-0.3%year^{-1} (IQR -0.6 to 0.1) (Fox *et al*., 2000b). Other studies have shown greater atrophy rates, possibly due to differences in cohorts or measurement method (Table 2-1). Indeed mean change in brain fractional volume over one year in 10 RRMS patients ranged from -1.6% to +2.1% depending on the acquisition and segmentation technique used (Leigh *et al*., 2002). In addition, with disease-modifying treatments now licensed for RRMS, it can be difficult to study the natural history of the disease in treatment-naïve subjects. To the best of my knowledge, the studies discussed in this section included subjects who were not on treatment at the time of study entry, and MRI acquisition was delayed for several days following the administration of steroids to treat relapses, due to the known effects that they have on brain volume (Hoogervorst *et al*., 2002; Rao *et al*., 2002). Additional examples of brain atrophy rates observed in RRMS patients can be seen in the retrospective analyses of placebo MRI data from treatment trials; atrophy rates in a range similar to those already detailed have been shown (Table 2-5). Lateral ventricular enlargement measures performed in RRMS patients are presented in Table 2-2.

Tissue specific atrophy has also been investigated in RRMS. Reductions in both the thickness and volume of the cortex have been shown in MS subjects relative to controls, with thinning observed in precentral, frontal, temporal and occipital regions early in the disease (Amato *et al*., 2004; Calabrese *et al*., 2007a; Sailer *et al*., 2003). Progressive cortical atrophy has also been observed in these areas (Chen *et al*., 2004). Ideally for any method of analysis, cross-sectional and longitudinal measures would be consistent, however SPM has shown no difference in GMF between 13 early RRMS subjects and controls at study entry but found a significantly greater change in GMF in patients over 18 months (Chard *et al*., 2004). The converse was found with WMF, which was significantly reduced in patients compared with controls at baseline, but there was no evidence that WMF changed at a different rate in the two groups. A recent extension of this study, including 21 subjects, supported the initial results except that baseline GMF was found to be significantly smaller in patients than controls (Tiberio *et al*., 2005). Similarly in a large study of 117 placebo subjects, mean estimated change in GM volume was -0.30% per month, whilst WM volume did not change significantly over the study period (Valsasina *et al*., 2005). Uncertainty over the precision and reliability of all these methods means that some of these results need to be interpreted cautiously.

Table 2-1 *Longitudinal studies of brain atrophy in relapsing remitting MS.*

Study	***MRI***	***Method***	***N***	***Age[a] (years)***	***Disease duration[a] (years)***	***EDSS (median (range) unless otherwise stated)***	***Estimated volume change per year[a]***	***Estimated volume change per year as a % of baseline brain volume[a]***
(Fox *et al*., 2000b)	Coronal T1-weighted, 1.5mm slices	Brain (BBSI)	6	36.7 (9.5)	5.6 (2.6)	3.25 (1.5-6.5)	-	median -0.8% (IQR -0.9 to 0.2)
(Ge *et al*., 2000b)	Axial dual echo, 3mm slices	Brain (FCP)	27	35.0 (5.2)	4.1 (2.3)	2.5 (SD 1.0)	median -17.3ml	-1.5%
(Kalkers *et al*., 2002)	Axial T1-weighted, 5mm slices (0.5mm interslice gap)	BPF (region growing)	42[b]	35.0 (8.6)	4.2 (4.5)	1.5 (IQR 1.0-2.0)	-	median -0.7% (IQR -1.3 to -0.01)
(Oreja-Guevara *et al*., 2005)	Axial T1-weighted 5mm slices, post-Gd-DTPA	Brain (SIENA)	26	36.0 (range 25-50)	10.0 (range 1-15)	1.5 (0-4.0)	-	-0.61% (SE 0.18%)
(Rovaris *et al*., 2000)	Axial T1-weighted, 3mm slices	Brain (region growing)	50	31.4 (7.3)	median 3 (range 1-13)	1.5 (0-4.0)	-14.9ml (range -65.1 to 38.7)	-1.3% (range -6.3 to 3.2)
(Rovaris *et al*., 2000)	Axial T1-weighted, 3mm slices	CCV (7 slices)	50	31.4 (7.3)	median 3 (range 1-13)	1.5 (0-4.0)	-5.3ml (range -22.3 to 5.7)	-1.7% (range -7.9 to 2.0)
(Rovaris *et al*., 2003)	Axial T1-weighted, 5mm slices	Brain (SIENA)	34	32.7 (8.4)	median 7 (range 2-25)	2.5 (1.0-5.5)	-	-0.7% (0.9)
(Sailer *et al*., 2001)	Axial T1-weighted, 5mm slices, post Gd-DTPA	CCV (4 slices)	13	38.2 (6.6)[c]	8.8 (6.3)[c]	6.0 (2-7)[c]	median -1.5cm^3 (range -20.7 to 2.1)	-0.51%[d]
(Saindane *et al*., 2000)	Axial dual echo, 3mm slices	Brain (FCP)	24	37.0 (7.5)	4.7 (3.3)	2.5 (SD 1.1)		-0.92% (1.20)

Continued from page 68

Study	***MRI***	***Method***	***N***	***Age[a] (years)***	***Disease duration[a] (years)***	***EDSS (median (range) unless otherwise stated)***	***Estimated volume change per year[a]***	***Estimated volume change per year as a % of baseline brain volume[a]***
(Tiberio *et al*., 2005)	Axial T1-weighted, 1.5mm slices	BPF (SPM)	21	37.5 (range 26.9-56.1)	2.1 (range 1.2-3.7)	1.0 (0-3.0)	-0.0058 (CI -0.0146 to -0.0055)	-0.75%
(Zivadinov *et al*., 2001b)	Axial T1-weighted, 5mm slices	Brain (Semi-auto Trieste method)	53	30.2 (9.4)	3.8 (1.3)	1.0 (0-5.0)	-16.2ml	-1.33%

[a]mean (SD) unless otherwise stated, [b]16 patients receiving interferon beta-1a at the time of follow-up examination, [c]includes data from 16 secondary progressive multiple sclerosis subjects also, [d]estimation based on a median baseline volume of 299.0cm^3, BBSI, brain boundary shift integral; BPF, brain parenchymal fraction; CCV, central cerebral volume; CI, 95% confidence interval; FCP, fuzzy connected principles; IQR, interquartile range; SE, standard error; SPM, statistical parametric mapping.

Table 2-2 *Longitudinal studies of lateral ventricular enlargement in relapsing remitting MS.*

Study	***MRI***	***N***	***Age[a] (years)***	***Disease duration[a] (years)***	***EDSS (median (range) unless otherwise stated)***	***Enlargement?***	***Estimated volume change per year***	***Estimated volume change per year as a percentage***
(Dalton *et al*., 2006)	Axial T1-weighted, 3mm slices	41	median 40 (range 22-62)	median 5 (range 1-25)	3.5 (2.0-7.0)	yes	median 0.5ml (range -1.7 to 4.2)	-
(Fox *et al*., 2000b)	Coronal T1-weighted, 1.5mm slices	6	36.7 (9.5)	5.6 (2.6)	3.25 (1.5-6.5)	yes	median 2.1ml (IQR 0.7 to 3.7)	-
(Kalkers *et al*., 2002)	Axial T1-weighted, 5mm slices (0.5mm interslice gap)	42[b]	35.0 (8.6)	4.2 (4.5)	1.5 (1.0-2.0)	yes	-	median 3.9%[c] (IQR 0.8 to 8.5)
(Luks *et al*., 2000)	Axial T1-weighted, 3mm slices	15	36 (range 18-56)	0.5 (range 2-18 months)	1.0 (0-4.0)	yes	mean 2.78cm^3	mean 20.2%[d] (range -5.5 to 91.1)
(Redmond *et al*., 2000)	Axial T1-weighted, 4mm slices, post Gd-DTPA	7	33 (7)	3.8 (1.7)[e]	mean 3.4 (range 1.5-6.0)	yes	median 4.6ml (IQR -1.88 to 5.02)	mean 29.8%[d]
(Turner *et al*., 2003)	Axial T1-weighted, 1mm slices, post Gd-DTPA	7	median 32 (range 21-47)[f]	median 4 (range 1-17)[f]	2.0 (0-4.0)[f]	yes	-	median 2.69%[d] (range 0.45-10.11)

[a]mean (SD) unless otherwise stated, [b]16 patients receiving interferon beta-1a at the time of follow-up examination, [c]as a percentage of baseline ventricular fraction (ventricular volume/intracranial volume), [d]as a percentage of baseline ventricular volume, [e]converted from months, [f]includes data from 13 patients receiving interferon beta-1a during the study.

2.3 Secondary progressive multiple sclerosis

Although smaller brain volumes have been shown in SPMS subjects compared with controls (Benedict *et al*., 2006; Bermel *et al*., 2003b; Lin *et al*., 2003; Turner *et al*., 2001), estimates of rates of tissue loss have varied considerably (Table 2-3 and Table 2-5). This may be due to the disease duration in these cohorts often being relatively heterogeneous. A mean -1.40% (SD 1.69) change in CCV was found in 38 SPMS placebo patients over the initial year of a 36 month study (Molyneux *et al*., 2000a). As discussed earlier (Chapter 1.6.3) this method is heavily weighted to ventricular enlargement and therefore may show higher rates of change than a whole brain measure. However whole brain techniques have shown similar annual rates of -1.4% (SD 2.0) (Rovaris *et al*., 2003) and -1.18% (SD 0.19) (Rovaris *et al*., 2005a). Whilst a decrease in brain volume of -2.0% $year^{-1}$ was observed in another analysis (Ge *et al*., 2000b), estimation may be less reliable due to the inclusion of only nine patients and varying follow-up periods (one to seven years). Atrophy measured on 3D volumetric images has been estimated at only -0.6% $year^{-1}$ (IQR -1.3 to 0.3) and -0.41% over 18 months (range -0.58 to -0.01) in two separate studies however (Fox *et al*., 2000b; Turner *et al*., 2003), although both estimations were based on only six subjects. Intermediate to these studies, a median -0.8% $year^{-1}$ (IQR -1.1 to -0.3) loss of brain tissue was observed in 21 patients (Kalkers *et al*., 2002). Estimates of lateral ventricular enlargement have varied and it is difficult to compare findings between studies. Over six months no enlargement was observed in one study; median change -0.07ml (IQR -0.82 to 0.66) (Redmond *et al*., 2000). However over 12 months in another study, median enlargement was 1.0ml (IQR 0.02 to 2.6) (Fox *et al*., 2000b). A change of 2.94% of baseline ventricular volume (range 0.56 to 19.73) was observed over 18 months in a further study (Turner *et al*., 2003). These rates were each estimated on only six or seven subjects however and disease duration ranged considerably between studies, so findings should be regarded with caution. A larger study of 23 subjects observed a statistically significant median increase of 1.1ml (range -1.1 to 6.9) (Dalton *et al*., 2006) which is the same as that observed in the study by Fox *et al.* (Fox *et al*., 2000b).

Although measures of global and cortical GM are observed to be lower in SPMS subjects than in controls (Benedict *et al*., 2006; Calabrese *et al*., 2007a; Carone *et al*., 2006), few studies have investigated longitudinal changes in this subgroup of patients so it is unclear whether significant progressive GM atrophy is occurring (Agosta *et al*., 2006; Chen *et al*., 2004).

Table 2-3 *Longitudinal studies of brain atrophy in secondary progressive MS.*

Study	***MRI***	***Method***	***N***	***Age[a] (years)***	***Disease duration[a] (years)***	***EDSS (median (range) unless otherwise stated)***	***Estimated volume change per year[a]***	***Estimated volume change per year as a % of baseline brain volume[a]***
(Fox *et al*., 2000b)	Coronal T1-weighted, 1.5mm slices	Brain (BBSI)	6	43.7 (6.3)	19.3 (2.9)	7.25 (6.0-8.0)	-	median -0.6% (IQR -1.3 to 0.3)
(Ge *et al*., 2000b)	Axial dual echo, 3mm slices	Brain (FCP)	9	46.3 (4.5)	5.4 (3.6)	mean 4.3 (SD 3.0)	median -23.6ml	-2.0%
(Kalkers *et al*., 2002)	Axial T1-weighted, 5mm slices (0.5mm interslice gap)	BPF (region growing)	21[b]	41.9 (8.8)	5.2 (7.2)	4.0 (2.5-5.5)	-	median -0.8% (IQR -1.1 to -0.3)
(Rovaris *et al*., 2003)	Axial T1-weighted, 5mm slices	Brain (SIENA)	19	40.5 (10.6)	median 8 (range 3-23)	5.5 (3.5-6.5)	-	-1.4% (2.0)
(Rovaris *et al*., 2005a)	Axial T1-weighted, 5mm slices	Brain (SIENA)	22[c]	48.3 (range 34-60)	median 17.9 (range 6-29)	6.0 (4.0-7.0)	-	-0.94% (0.15)
(Sailer *et al*., 2001)	Axial T1-weighted, 5mm slices, post Gd-DTPA	CCV (4 slices)	16	38.2 (6.6)[d]	8.8 (6.3)[d]	6.0 (2-7)[d]	median -3.0cm^3 (range -8.3 to 2.1)	-1.01%[e]

[a]mean (SD) unless otherwise stated, [b]six patients receiving interferon beta-1a at the time of follow-up examination, [c]15 patients treated with disease-modifying treatments during follow-up, [d]includes data from 13 relapsing remitting multiple sclerosis subjects, [e]estimation based on a median baseline volume of 299.0cm^3, BPF, brain parenchymal fraction; CCV, central cerebral volume; FCP, fuzzy connected principles; IQR, interquartile range.

2.4 Primary progressive multiple sclerosis

Like SPMS fewer longitudinal studies investigating atrophy in PPMS have been performed, although cross-sectional analysis has shown reduced brain volumes in PPMS compared with controls (De Stefano *et al*., 2003; Sastre-Garriga *et al*., 2004). One large longitudinal analysis of 137 PPMS patients estimated a change in CCV of -1.3% $year^{-1}$ (Stevenson *et al*., 2000) (Table 2-4). A subset of these patients (n=100) were followed up for a second year and by the end of this period a -2.7% change relative to baseline was observed (Ingle *et al*., 2002). Using the same method in another 39 subjects followed for 18-28 months the mean change was -2.31% (SD 2.69) (Stevenson *et al*., 2002). Interestingly atrophy was quantified using SIENA in these same 39 subjects and was estimated to be only -0.56% (SD 0.57). This disparity may be because the two techniques measure different volumes, and volume loss is greater around the ventricles; the CCV change may be driven strongly by the ventricular expansion, thereby giving higher rates of change but, equally, higher variances. Brain atrophy rates in three other studies using SIENA ranged from -0.64 to -1.24% over approximately 12 months, which did not appear to be related to disease duration (Jasperse *et al*., 2007a; Rovaris *et al*., 2005a; Sastre-Garriga *et al*., 2005a). Estimated brain atrophy rate on 3D FSPGR images was -0.9% $year^{-1}$ (IQR -1.4 to -0.3) using the BBSI (Fox *et al*., 2000b) and -1.03% $year^{-1}$ (SD 1.30) using BPF (SPM method) (Sastre-Garriga *et al*., 2005a). Progressive brain atrophy and ventricular enlargement has been observed over periods up to five years (Ingle *et al*., 2003). Lateral ventricular enlargement of 1.68ml $year^{-1}$ (SD 11.3) and 2.9ml $year^{-1}$ (IQR 0.6 to 4.5) has been observed (Fox *et al*., 2000b; Stevenson *et al*., 2004), but over a shorter period of six months no significant atrophy appeared to occur (mean -0.24ml (IQR -1.84 to 0.71) (Redmond *et al*., 2000); however only five patients were included in this analysis and the study lacked power.

As in RRMS and SPMS cohorts, GM atrophy appears to progress at a relatively greater rate than WM atrophy: -1.50% (SD 1.6%) compared with -0.07% in one investigation (Sastre-Garriga *et al*., 2005a).

Table 2-4 *Longitudinal studies of brain atrophy in primary progressive MS.*

Study	***MRI***	***Method***	***N***	***Age[a] (years)***	***Disease duration[a] (years)***	***EDSS (median (range) unless otherwise stated)***	***Estimated volume change per year[a]***	***Estimated volume change per year as a % of baseline brain volume[a]***
(Fox *et al*., 2000b)	Coronal T1-weighted, 1.5mm slices	Brain (BBSI)	9	49.3 (8.7)	11.3 (5.8)	6.0 (3.5-8.5)	-	median -0.9% (IQR -1.4 to -0.3)
(Jasperse *et al*., 2007a)	Axial T1-weighted, 5mm slices	Brain (SIENA)	15	43.6 (8.9)	median 1.5 (IQR 0.9-3)	3.0 (2.5-4.0)	-	-0.9% (0.6)
(Kalkers *et al*., 2002)	Axial T1-weighted, 5mm slices (0.5mm gap)	BPF (region growing)	20	47.9 (10.3)	7.1 (8.1)	4.0 (2.5-6.0)	-	median -0.5% (IQR -0.9 to -0.2)
(Rovaris *et al*., 2005a)	Axial T1-weighted, 5mm slices	Brain (SIENA)	54	51.3 (range 25-68)	median 10 (range 2-26)	5.5 (2.5-7.5)	-	-0.99% (0.13)
(Sastre-Garriga *et al*., 2005a)	Axial T1-weighted, 1.5mm slices	Brain (SIENA)	31	median 46 (range 26-62)	median 3.0 (range 2-5)	4.5 (3.5-7.0)	-	-0.63% (1.05)
(Sastre-Garriga *et al*., 2005a)	Axial T1-weighted, 1.5mm slices	BPF (SPM)	31	median 46 (range 26-62)	median 3.0 (range 2-5)	4.5 (3.5-7.0)	-	-1.03% (1.30)
(Stevenson *et al*., 2000)	Axial T1-weighted, 3mm slices	CCV	137[b]	not given	not given	6.0 (2.0-8.5)	-3.44ml (6.85)	-1.30%
(Stevenson *et al*., 2002)	Axial T1-weighted, 3mm slices	CCV	39	51 (range 29-74)	10.7 (range 1-26)	6.0 (2.0-8.5)	-	-1.22% (1.42)
(Stevenson *et al*., 2002)	Axial T1-weighted, 3mm slices	Brain (SIENA)	39	51 (range 29-74)	10.7 (range 1-26)	6.0 (2.0-8.5)	-	-0.29% (0.30)

[a]mean (SD) unless otherwise stated, [b]16 patients receiving disease-modifying treatment, BBSI, brain boundary shift integral; BPF, brain parenchymal fraction; CCV, central cerebral volume; FCP, fuzzy connected principles; IQR, interquartile range.

2.5 Regional atrophy

Volumetric measurements of regional structures thought to be differentially affected in MS may prove more sensitive markers of pathology than whole brain atrophy. Group analysis of parenchymal volume in 13 different hemispheric regions in RRMS and SPMS subjects found that the largest reductions in brain volume relative to controls were in the GM areas of the posterior basal ganglia/thalamus, superior frontal and superior parietal regions (Carone *et al*., 2006). Mean percentage volume differences relative to control subjects were -19.3%, -15.7% and -14.3% respectively in these regions. Similarly in RRMS a VBM study found GM was significantly reduced in the left frontotemporal cortex, anterior cingulate gyrus and bilateral caudate nuclei (Prinster *et al*., 2006). In concordance with these findings bicaudate ratio (BCR), a linear atrophy measure, has been shown to be greater in MS patients than controls (Bermel *et al*., 2002). Although this measure could represent whole brain atrophy and subsequent ventricular enlargement, caudate volumes when measured directly have been reduced, by 19% in RRMS subjects compared with controls in one study (Bermel *et al*., 2003a).

The thalamus may be a good structure through which to investigate atrophy, specifically GM atrophy. It has well defined boundaries, thereby minimising partial volume effects, and has extensive reciprocal cortical and subcortical connections and may therefore be sensitive to effects of pathology in widespread areas. VBM analysis has also identified progressive thalamic atrophy in RRMS and PPMS subjects when compared with controls (Audoin *et al*., 2006; Sepulcre *et al*., 2006). A post-mortem study of 10 subjects (one RRMS, six SPMS and three PPMS) found a 22% decrease in whole thalamic volume compared with controls, and decreased neuronal density (Cifelli *et al*., 2002). The same study obtained normalised thalamic volumes after manual outlining on 3D FSPGR images in a further 14 SPMS subjects, and a 17% decrease was observed relative to controls. Likewise, in 14 RRMS subjects, normalised thalamic volume was 25% less than controls, and volume was correlated with disease duration (Wylezinska *et al*., 2003). However a longitudinal study of eight RRMS and three SPMS patients found mixed results (Taylor *et al*., 2004). Average (of three measurements by the same operator) left and right percentage thalamic volume change varied between -8.5% and +14.2%, whilst whole brain atrophy was detected in all but one of the subjects. The range was partly due to differing study periods (ranging from 12 to 41 months) but the increased thalamic volumes in six of the subjects may be due to the difficulty of reliably outlining thalamic

regions on T2-weighted images compared with 3D volumetric T1-weighted coronal images.

2.6 Clinical and research applications of brain atrophy measures

2.6.1 Brain atrophy as a marker of disability

The key factors when assessing atrophy measurements are a) how well they relate to progression of disability, and b) whether they explain disability better than established measures of focal inflammation. Although redistribution of Na^+ channels on axons and cortical reorganisation have been observed in MS and may result in a return of function (Craner *et al*., 2004; Rocca *et al*., 2003), these processes are likely to become less effective over time as a threshold of damage is reached, beyond which it is impossible to compensate for the changes occurring. Evidence of significantly greater brain atrophy rates in CIS subjects who develop MS compared with those who remain clinically stable has been presented (Brex *et al*., 2000; Dalton *et al*., 2002a; Dalton *et al*., 2004; Filippi *et al*., 2004). However in a group of CIS subjects, no significant correlation between annual ventricular enlargement and baseline EDSS, one year EDSS or the change in EDSS was shown (Dalton *et al*., 2002a). It may be that atrophy in the early stages of MS does not result immediately in clinical disability (possibly due to cortical reorganisation), and that there is a threshold of axonal loss, which may vary between individuals, beyond which progressive disability is apparent. Equally, measurement of volume loss may be confounded by active inflammation resulting in some areas of tissue swelling. In MS subjects showing sustained progression in disability on the EDSS, significantly greater atrophy rates have been observed compared with subjects showing no progression in disability (Coles *et al*., 1999; Ingle *et al*., 2002; Jasperse *et al*., 2007a; Losseff *et al*., 1996; Molyneux *et al*., 2000a; Rudick *et al*., 2000; Zivadinov *et al*., 2001b). Equally, subjects who have shown a significant amount of brain atrophy over periods of up to four years are more likely to show an increase in EDSS score than subjects without significant atrophy (Gasperini *et al*., 2002; Losseff *et al*., 1996; Turner *et al*., 2003). GM atrophy may be particularly relevant; progressive thalamic atrophy has been shown to be related to EDSS change over two years (Audoin *et al*., 2006) and significantly greater cortical thinning has been observed in patients with progressive disability compared with those that are stable (Chen *et al*., 2004). Although some investigations have shown no relationship of atrophy with disability (Fox *et al*., 2000b; Kalkers *et al*., 2002; Rovaris *et al*., 2001; Stevenson *et al*., 2004), this may be due to the insensitivity of clinical disability

scales or that atrophy may not cause immediate effects. Significant correlations between change on clinical disability scales and brain atrophy have also been observed over periods of two to six years (Fisher *et al*., 2000; Rudick *et al*., 2001; Zivadinov *et al*., 2001b). Importantly, an eight year follow-up of 138 RRMS subjects who had taken part in a trial of interferon beta-1a, showed that atrophy rate during the two year trial was the only significant MRI predictor of disability status at year eight (Fisher *et al*., 2002). Similarly brain atrophy over two years was significantly correlated with EDSS at five year follow-up in another study (Rovaris *et al*., 2007). It appears that atrophy is a relevant marker of disease progression and may precede the development of measurable disability, but the mechanisms of this interaction need to be investigated further. It is likely that there is heterogeneity between subjects, which is evidenced by the fact that similar atrophy rates have been observed in RRMS and PPMS, but that disability progresses differently.

2.6.2 Brain atrophy as a marker of cognitive impairment

Cognitive impairment is seen in around 50% of MS patients even early in the disease and appears to progress over time (Achiron & Barak, 2003; Amato *et al*., 2001; Lazeron *et al*., 2005; Zivadinov *et al*., 2001b). Typically deficits in processing speed and attention are observed, although a range of subcortical and cortical deficits may be seen, such as short- and long-term verbal and visuospatial memory, which may affect daily living for people with MS (Amato *et al*., 2001; Lazeron *et al*., 2006; Sanfilipo *et al*., 2006). Brain atrophy may be useful as a marker or predictor of cognitive changes and it has been shown to be associated with neuropsychological impairment where correlations with MRI lesion measures have been weak or absent (Lazeron *et al*., 2006; Sanfilipo *et al*., 2006; Zivadinov *et al*., 2001b). Significantly smaller normalised brain volumes (NBV) have been observed in RRMS subjects considered to be impaired on neuropsychological tests of verbal memory, verbal fluency and attention/concentration compared with unimpaired patients (Amato *et al*., 2004; Zivadinov *et al*., 2001a). In SPMS and PPMS also, subjects have been shown to perform poorly in tests of verbal memory, verbal fluency, attention and spatial reasoning, and composite cognitive impairment scores have been shown to correlate significantly with cerebral volume (Camp *et al*., 1999; Molyneux *et al*., 2000a). Correlations between longitudinal measures of brain atrophy rate and cognition have also been observed; in both early RRMS and SPMS, subjects who worsened on neuropsychological tests had greater brain atrophy rates than the subjects whose cognition was stable or improved (Molyneux *et al*., 2000a; Zivadinov *et al*., 2001b).

Whilst regional atrophy measurements may not be an accurate indicator of global cognitive changes, the size of the CC in MS has been found to correlate with measures of mental processing speed and rapid problem solving (Pelletier *et al*., 2001). However one must be careful in ascribing a direct structure-function association in these types of studies, where only one region is measured. In addition, superior frontal cortex atrophy has been correlated with tests of new learning, divided attention and conceptual reasoning (Benedict *et al*., 2002). Cortical volume has also been found to be reduced in mildly cognitively impaired RRMS subjects compared with cognitively normal RRMS subjects (Amato *et al*., 2004; Benedict *et al*., 2006; Portaccio *et al*., 2006). A VBM study found that a test of processing speed/working memory was correlated not only with global GM volume, but with GM volume in regions associated with working memory and executive function including the bilateral prefrontal cortex (Morgen *et al*., 2006). In concordance, measures of regional GM have been associated with neuropsychological impairment; left frontal atrophy was associated with auditory and verbal memory tests whilst right frontal atrophy was associated with visual episodic and working memory (Tekok-Kilic *et al*., 2007). Regional atrophy measures of GM and WM may aid in our understanding of specific cognitive changes occurring throughout the disease. More longitudinal studies need to be performed, but may be difficult due to the relative changes in cognition being small, changes to imaging and scanners over these periods and practice effects on neuropsychological tests.

2.6.3 Prognosis for patients with clinically isolated syndromes and early multiple sclerosis

Brain atrophy rate could prove to be a useful marker for future prognosis in CIS or early MS patients in terms of MS diagnosis, disability and speed of progression, and cognitive decline. Studies described in Chapter 2.1 show that CIS subjects who go on to develop MS have significantly greater atrophy rates at the earliest stages of disease. In addition, early brain atrophy rate may be related to cognitive changes that can influence daily living. Longer follow-up studies of brain atrophy and clinical characteristics in these patient cohorts are required.

2.6.4 Understanding the causes of multiple sclerosis

Apolipoprotein E (ApoE) expressed in the brain plays an important role in lipid and cholesterol transport and is involved in growth and regeneration of neurons. Experimental

and clinical evidence exists for a pathogenic role of ApoE in Alzheimer's disease (AD), the prototypic neurodegenerative disease, but its role has also been investigated with respect to MS. Of 76 RRMS subjects in one study, 18 had the ApoE e4 allele and had significantly smaller NBV compared with controls and non-carriers, who were matched in age and disease duration (De Stefano *et al*., 2004). The same result was found when only subjects with short disease duration (less than three years) and low disability (EDSS < 2) were studied, suggesting that the e4 allele may either affect brain development or brain integrity very early in the disease. In contrast, a study that determined ApoE genotype in 117 MS patients and 100 controls, found that it was not related to the degree of brain atrophy (Zakrzewska-Pniewska *et al*., 2004), although atrophy was assessed using a subjective visual rating scale, which may have led to inaccuracy.

Longitudinal studies have also produced contrasting results. In RRMS patients who underwent ApoE genotyping, significant differences were found in brain atrophy rates between six different genotype groups, mainly due to patients possessing the e4 allele (Enzinger *et al*., 2004). Brain volume loss was -0.66% $year^{-1}$ (SD 0.62) in 22 subjects with one or two e4 alleles but only -0.13% $year^{-1}$ (SD 0.36) in 76 subjects without an e4 allele, despite longer disease duration. A separate study of 174 MS patients found no major association of ApoE genotype and brain atrophy rate however (Zwemmer *et al*., 2004).

2.6.5 Understanding disease mechanisms

Relationship to lesions

There is evidence that focal inflammation may lead to brain atrophy, but correlations of brain atrophy with lesion measures have been mixed. Whilst some studies of CIS and MS have shown an association between brain volumes and T1-hypointense lesions (Chard *et al*., 2002b; Dalton *et al*., 2004; Sailer *et al*., 2003), T2-hyperintense lesions (Chard *et al*., 2002b; Dalton *et al*., 2004; De Stefano *et al*., 2003; Lin *et al*., 2003; Sailer *et al*., 2003; Sastre-Garriga *et al*., 2004) and Gd-enhancing lesions (Lin & Blumhardt, 2001; Luks *et al*., 2000), others have not (Chard *et al*., 2002b; De Stefano *et al*., 2003; Ge *et al*., 2000b). However cross-sectional studies present only an isolated view of ongoing disease activity and comprehensive evaluation of the relationship requires assessment of serial data.

Significantly greater ventricular enlargement has been observed over one year in CIS subjects with T1, T2 or Gd-enhancing lesions at baseline, compared with subjects who had no lesions on baseline imaging (Dalton *et al*., 2002a). Moreover, this study suggested that atrophy was to a greater extent due to the presence of T1 lesions than T2 lesions. Likewise studies in MS have found the number and volume of lesions at baseline or over the initial study phase are significantly correlated with brain atrophy over periods of 18 months to three years (Gasperini *et al*., 2002; Jasperse *et al*., 2007a; Luks *et al*., 2000; Molyneux *et al*., 2000a; Paolillo *et al*., 2004). Several studies have shown a relationship between brain atrophy and change over the same period in lesion volumes (Horakova *et al*., 2007; Ingle *et al*., 2002; Richert *et al*., 2006; Rudick *et al*., 1999; Rudick *et al*., 2000), however an absence of any correlation has also been observed (Gasperini *et al*., 2002; Losseff *et al*., 1996; Rashid *et al*., 2007; Rovaris *et al*., 2001; Rudick *et al*., 1999; Stevenson *et al*., 2002). It may be that focal inflammation has a delayed effect on neuroaxonal degeneration and subsequent atrophy. Several studies including longer patient follow-up support this argument. One study following 28 RRMS subjects for 14 years after first symptoms found that the change in lesion load in the first five years after onset was more closely correlated to brain atrophy at 14 years than later changes in lesion load (Chard *et al*., 2003). Similarly a 13 year follow-up of RRMS patients found that the change in T2 lesion volume during the first two years correlated significantly with BPF at year 13 (Rudick *et al*., 2006). In two-year trials of interferon beta-1a greater brain atrophy at six and eight year follow-up respectively was significantly associated with Gd-enhancing, T1 and T2 lesion loads during the pre-trial and trial periods (Fisher *et al*., 2002; Paolillo *et al*., 2002).

The absence of a correlation between lesions and atrophy seen in some studies and the only moderate correlations seen in others could be due to factors such as small lesion loads, lesion activity resulting in different degrees of axonal damage, inflammation causing oedema which masks atrophy, or GM lesions (not commonly seen on MRI) having a greater effect on subsequent atrophy than WM lesions. Little research has investigated the association of tissue fractions or regional atrophy with global or regional lesion measures due to technical difficulties. However there is evidence that factors unrelated to lesion formation may play a role in atrophy progression. PPMS have fewer lesions than patients experiencing relapses, but progressive atrophy occurs (Ingle *et al*., 2003; Sastre-Garriga *et al*., 2004). In addition absence of a correlation between lesion

measures and WMF has been observed in at least two studies (Chard *et al*., 2002b; Dalton *et al*., 2004) whilst a study showed that administration of autologous haematopoietic stem cells had a dramatic and sustained effect on Gd-enhancement and T2 lesion formation but did not inhibit atrophy (-1.9% $year^{-1}$) (Inglese *et al*., 2004). A similar effect has been observed with Campath-1H, a powerful lymphocyte depleting monoclonal antibody (Coles *et al*., 1999). It appears that whilst possibly conditioned by inflammatory lesion load, atrophy may proceed even in the absence of evidence of inflammation, and other mechanisms underlying neuroaxonal degeneration must be explored. Once again, different time courses for these processes may confound studies of short duration: for instance if there is a long delay between the effects of lesions (or treatment) and the progression of atrophy.

Progression of atrophy

Atrophy appears to occur from the earliest stages of MS and continues into the progressive stages of disease, but it is unclear whether differences in the development of atrophy exist between progressive and relapsing forms of MS, or if atrophy rate changes during the course of disease. Investigation is difficult also because of treatment intervention in some patients and groups of patients. One study has suggested that atrophy is confined to the cerebral hemispheres during the RR stage, but extends to the cerebellum, brainstem and spinal cord during the SP phase (Lin *et al*., 2003). A trend for global cortical thickness to be lower in subjects with SPMS relative to subjects with RRMS has been observed (Calabrese *et al*., 2007a). Recently, SPM analysis of brain atrophy measured by SIENA was used to investigate the evolution of brain atrophy in MS patients according to phenotype (RRMS, SPMS and PPMS) (Pagani *et al*., 2005b). Results suggested that ventricular enlargement was predominant in RRMS, whilst cortical atrophy was more important in progressive MS.

Cross-sectional associations between normalised brain volume and disease duration (Bermel *et al*., 2003b; Ge *et al*., 2000b; Kalkers *et al*., 2001a; Paolillo *et al*., 2000; Rudick *et al*., 2000) suggest brain atrophy is progressive, and a longitudinal analysis of 27 RRMS and nine SPMS patients, found that brain atrophy rate was greater in patients with SPMS, who had a longer disease duration, than in patients with RRMS (Ge *et al*., 2000b). In addition no significant differences in NBV or annual atrophy rate have been observed in two studies of patients with RRMS and PPMS where the subjects had a similar disease

duration (De Stefano *et al*., 2003; Jasperse *et al*., 2007a). However, a trend for greater annual atrophy rates in RRMS and SPMS compared with PPMS has also been observed (Kalkers *et al*., 2002), despite PPMS subjects having longer disease duration. Whether atrophy rates decelerate or accelerate during the course of progressive disease is unclear. Analysis of 100 subjects with PPMS found that the degree of atrophy over the first year did not correlate with that over the second year (Ingle *et al*., 2002). However a subsequent five year follow-up report, found a relatively consistent atrophy rate within individuals (Ingle *et al*., 2003). In addition, no obvious change in the atrophy rate over 36 months was observed in 44 placebo subjects with SPMS taking part in a trial of interferon beta-1a who were observed at six-monthly intervals (Molyneux *et al*., 2000a), or in a five year study of 36 patients on combination therapy analysed annually (Horakova *et al*., 2007). Studies of patients over longer periods need to be performed in order to clarify the temporal dynamics of atrophy in MS. However this may be difficult due to patient drop-out from studies and changes in MR imaging and scanner upgrades.

2.6.6 Clinical trials and sample size calculations

Although there is no cure for MS, several disease-modifying treatments have been tested in patients, some of which are now licensed for administration in RRMS and SPMS following demonstration of their capacity to reduce relapse rates and slow the progression of mild clinical disability (Jacobs *et al*., 1996; Johnson *et al*., 1998; The IFNB Multiple Sclerosis Study Group, 1995). Brain atrophy measurements have been applied retrospectively to data from some of these clinical trials, to assess their effect on neurodegeneration, and more recently it has been included as a secondary (and retrospective) outcome measure in treatment trials (Miller *et al*., 2007; Rovaris *et al*., 2001; Rudick *et al*., 1999) (Table 2-5). Brain atrophy appears to continue despite treatment with these putative disease-modifying drugs, although longer follow-up periods in three studies suggest that either therapeutic action is delayed, or that a beneficial effect from baseline becomes apparent only in later atrophy measures (Frank *et al*., 2004; Paolillo *et al*., 2002; Turner *et al*., 2003).

As chronic disease and disability progression is the principle clinical challenge, and development of neuroprotective agents is a key objective, brain atrophy rate as an outcome measure is becoming increasingly important. It is vital that clinical trials attempting to measure cerebral atrophy as a marker of progression are of sufficient power

to detect a treatment effect. Determining natural brain atrophy rates in MS allows some estimation of the patient numbers that are required if clinical trials are to show a treatment effect. Large variability in atrophy rates in patient and control groups, whether due to measurement error or biological inter-subject differences, will increase the sample sizes required in parallel group design trials however. Calculation of sample sizes from different brain atrophy measurement techniques will help to identify the best method to implement into future trials of disease-modifying drugs. Evidence that different acquisitions and measurement techniques affect observed atrophy rates advocates standardisation of acquisition and methods to increase the reproducibility and robustness of results, especially in the context of multicentre longitudinal clinical trials (Leigh *et al*., 2002).

If brain atrophy is to be used as a marker of neuronal or axonal degeneration in clinical trials, reliable detection of atrophy over short intervals would potentially improve trial efficiency and reduce costs. However the effect of measurement errors on brain atrophy quantification is increased over short intervals, and consequently results in greater variance of atrophy rate measures. Sufficient power to detect significant decreases in brain volume over short intervals is therefore likely to require much larger subject numbers than would be expected in a longer interval study. Over short intervals, results of atrophy studies have been mixed. In two studies of 30 and 28 patients with RRMS respectively, it was not possible to show significant brain atrophy over three months using three different BPF calculation methods or SIENA (Fritz *et al*., 2006; Zivadinov *et al*., 2004a). Nonetheless, a significant mean change in BPF of -0.23% was observed over approximately 2.5 months in 128 patients in one study, and estimated annual atrophy rate was -1.06% (95% CI -1.50% to -0.62%), a figure comparable to studies following patients over one year or more (Hardmeier *et al*., 2003). One must be careful extrapolating results obtained over such short intervals however due to the fact that measured atrophy rates are unlikely to be linear.

Table 2-5 *Brain atrophy in MS cohorts taking part in therapeutic trials.*

Study	***Treatment***	***Method***	***Group***	***N (placebo/treated)***	***MRI assessments (months)***	***Estimated atrophy in placebo over interval[a] (shown in months)***	***Estimated atrophy in treated over interval[a] (shown in months)***	***Significant difference in atrophy rate, placebo vs treatment?***	***Significant decrease in atrophy rate in treated?***
(Filippi *et al*., 2004)	IFN β-1a (ETOMS study)	Brain (SIENA)	CIS	98/111	12/24	0-12=-0.83% (1.09) 12-24=-0.67% (1.10)	0-12=-0.62% (1.40) 12-24=-0.61% (0.99)	0-24 placebo>treated	-
(Frank *et al*., 2004)	IFN β-1b	Brain	RRMS	0/30	12/24/36	-	0-12=-1.35% (0.79) 0-24=-1.48% (0.77) 0-36=-1.68% (0.73)	-	yes
(Gasperini *et al*., 2002)	IFN β-1a	CCV	RRMS	0/52	monthly -6-9/ 12/24	-6-0=0.02%	6-30=-2.2%	-	no
(Ge *et al*., 2000a)	Glatiramer acetate (US study)	Brain (FCP)	RRMS	13/14	24	0-24=-1.8% year^{-1} (1.8)	0-24=-0.6% year^{-1} (0.9)	0-24 placebo>treated	-
(Hardmeier *et al*., 2005)	IFN β-1a (European study)	BPF (Cleveland)	RRMS	0/386	12/24/36	-	0-12=-0.69% (0.79) 12-24=-0.38% (0.77) 24-36=-0.38% (0.73)	-	yes
(Miller *et al*., 2007)	Natalizumab (AFFIRM study)	BPF (Cleveland)	RRMS	315/627	12/24	0-12=-0.40% 12-24=-0.43%	0-12=-0.56% 12-24=-0.24%	0-12 treated>placebo 12-24 placebo>treated	-
(Molyneux *et al*., 2000a)	IFN β-1b (European study)	CCV	SPMS	46/49	6/12/18/24/30/ 36	0-6=-0.89% (1.35) 0-12=-1.40% (1.69) 0-18=-2.02% (2.83) 0-24=-2.76% (3.30) 0-30=-2.74% (3.38) 0-36=-3.86% (3.53)	0-6=-1.39% (1.47) 0-12=-1.60% (2.50) 0-18=-1.65% (2.34) 0-24=-2.17% (2.90) 0-30=-2.84% (3.12) 0-36=-2.91% (3.11)	no	no

Continued from page 84

(Rovaris *et al*., 2001)	Glatiramer acetate (Euro/Canadian study)	CCV	RRMS	114/113 (9-18 months open-label)	9/18	0-9=-0.7% (2.2) 9-18=-0.6% (2.0)	0-9=-0.8% (1.9) 9-18=-0.4% (1.7)	no	no
(Rovaris *et al*., 2007)	Glatiramer acetate (Euro/Canadian study) – long-term follow-up	Brain (SIENA)	RRMS	69/73	60	18-60=-3.46% (2.47)	18-60=-3.32% (2.40)	no	-
(Rudick *et al*., 1999)	IFN β-1a	BPF (Cleveland)	RRMS	72/68	12/24	0-12=-0.70% (0.92) 12-24=-0.52% (0.80)	0-12=-0.76% (1.11) 12-24=-0.23% (0.74)	12-24 placebo>treated	-
(Sormani *et al*., 2004)	Glatiramer acetate (Euro/Canadian study)	Brain (SIENA)	RRMS	105/102 (9-18 months open-label)	9/18	0-9=-0.9% (1.2) 9-18=-1.0% (1.1)	0-9=-0.8% (1.0) 9-18=-0.6% (1.2)	9-18 placebo>treated	-
(Turner *et al*., 2003)	IFN β-1a (PRISMS study)	Brain (Seg prop)	RRMS	7/13	18/48	0-18=-1.11% (range -2.25 to 1.10) 0-48=0.83% (range -3.47 to 1.24)	0-18=-0.69% (range -2.65 to 0.08) 0-48=-1.02% (range -7.52 to 0.64)	0-18 placebo>treated	no
(Turner *et al*., 2003)	IFN β-1a (SPECTRIMS study)	Brain (Seg prop)	SPMS	6/12	18/48	0-18=-0.41% (range -0.58 to -0.01) 0-48=-0.26% (range -1.16 to 1.34)	0-18=-0.51% (range -3.03 to 1.67) 0-48=-0.32% (range -2.90 to 1.05)	no	yes
(Zivadinov *et al*., 2007)	IFN β-1a	BPF (SPM)	RRMS	28/26	36	BPF =-2.5% GMF=-1.4%	BPF=-1.3% GMF= 0.2%	yes	-

[a]*mean (SD) unless otherwise stated, BPF, brain parechymal fraction; CCV, central cerebral volume; FCP, fuzzy connected principles; IFN, interferon; Seg prop, segmentation propagation.*

2.7 Chapter conclusions

In conclusion, MRI measures of brain atrophy provide a feasible *in vivo* measure of neuroaxonal degeneration in MS. A wide range of methods have been developed to measure global atrophy that have been shown to be sensitive to changes from early in the course of disease to later progressive stages. Brain atrophy rate appears to provide a better marker for clinical disability than conventional lesion measures and may provide clues for patient prognosis and understanding disease mechanisms. As the clinical applications of brain atrophy are moving forward and the likelihood that it will be applied as an outcome measure in clinical trials increase, more work validating different brain atrophy measures in MS and determining the best measure is required. In addition to improving the stability of MR acquisitions, easily applicable methods for quantifying brain atrophy and improved measurement precision, will be particularly important factors with the future emergence of putative neuroprotective agents.

However there are important discrepancies between brain atrophy and clinical measures. It is still unclear when and at what rate atrophy develops, how it relates to pathology, and whether a threshold of atrophy exists beyond which progressive disability is inevitable. The location of atrophy, and by implication axonal loss, will also be relevant. Longitudinal MRI and clinical studies of subjects from first symptoms to post-mortem are required to give insight into the underlying disease mechanisms and verify MRI findings.

3 Methods overview

3.1 Subjects

Patients and control subjects used in the studies described in this thesis were taking part in two longitudinal clinical and MRI research studies being undertaken at the Institute of Neurology, University College London, Queen Square, London, United Kingdom. The aim of these studies was to investigate early MR markers of prognosis and pathogenic mechanisms in MS and details of the cohorts are described below. The studies in this thesis were a retrospective analysis of the MR and clinical data obtained. Ethical approval for the studies was obtained from the Joint Medical Ethics Committee of the Institute of Neurology and National Hospital for Neurology and Neurosurgery, University College London, Queen Square, London, United Kingdom. In addition, written informed consent was obtained from all study participants. All patients underwent appropriate clinical and laboratory investigations to exclude alternative diagnoses.

3.1.1 Clinically isolated syndrome study

Approximately 90% of patients were recruited to this study from Moorfields Eye Hospital where they had presented with optic neuritis. Other patients were recruited from general neurology and out-patient clinics at the National Hospital for Neurology and Neurosurgery where they had presented with a CIS suggestive of MS including brainstem and spinal cord syndromes. A baseline assessment for the study was performed within 12 weeks of presentation with a CIS. Further clinical and MRI assessments were performed at three months, one year, three years and five years. A subsequent diagnosis of MS in CIS patients was made according to the McDonald criteria (McDonald *et al*., 2001).

3.1.2 Relapsing remitting multiple sclerosis study

Patients were recruited from general neurology and out-patient clinics at the National Hospital for Neurology and Neurosurgery. All patients had clinically definite MS according to the Poser criteria (Poser *et al*., 1983) and fulfilled established criteria for RRMS (Lublin & Reingold, 1996). A total of 41 patients were recruited within four years of symptom onset and none had previously received disease-modifying treatment. Patients on disease-modifying treatment or with longer disease duration were excluded from the study. None of the patients had experienced a clinical relapse or received

corticosteroids within the month before assessments. Patients were assessed at baseline and then at six-monthly intervals for up to three years. An additional assessment was performed at five years, which included cognitive assessment.

3.1.3 Control subjects

Control subjects were all healthy volunteers with no previous history of neurological or major medical disease, and three control cohorts were utilised in this study. All subjects gave written informed consent for involvement in imaging studies. Twenty-six subjects were recruited as controls in the longitudinal study of RRMS. Assessment was at baseline with follow-up at six-monthly intervals for up to three years. These subjects were staff, or relatives and friends of staff, from within the National Hospital for Neurology and Neurosurgery and Institute of Child Health, or were relatives and friends of patients taking part in the longitudinal study of RRMS.

In addition, five control subjects, who were staff at the Institute of Neurology, were recruited to be scanned over a planned MRI scanner upgrade. These subjects were scanned approximately 10 times over a 24 week period prior to the upgrade and approximately 10 times over a 33 week period following the upgrade.

A further 22 control subjects who had two MR scans acquired on the same day were also used in this thesis. These subjects were taking part in a project aimed at determining the shortest interval required to detect volumetric change based on MRI, to distinguish AD subjects from normal controls (The Minimum Interval Resonance Imaging in AD (MIRIAD) project). In addition to two MR scans being acquired at baseline, MRI was acquired at two, six, 14, 26, 38 and 52 weeks.

3.2 Clinical assessment of patients

At baseline assessment a comprehensive patient history and neurological examination was performed. At subsequent assessments a history since the last assessment was taken, including any relapses (number, time and length of relapse, symptoms), a review of current medication (if any) or changes in medication, and any other pathology. Neurological examination included assessment and scoring of Kurtzke's Functional Systems and EDSS (Kurtzke, 1983) through a combination of questions and physical assessments by the examiner (Appendix 1). The Multiple Sclerosis Functional Composite

(MSFC) (Appendix 2) (Cutter *et al*., 1999; Fischer *et al*., 1999), a composite measure of clinical impairment, was assessed in patients with RRMS at all time-points. This encompassed a test of cognition (the paced auditory serial addition test (PASAT)), a test of upper limb function (the nine-hole peg test (9HPT)) and a test of lower limb function (the 25-foot timed walk (TW)).

3.3 Magnetic resonance imaging

All brain MRI was acquired on a 1.5 Tesla GE Signa Horizon Echospeed scanner (General Electric Medical Systems, Milwaukee, WI, USA) on the same day as clinical assessment. The details of the acquisitions are described below. Image analysis was performed on digitised images.

3.3.1 Clinically isolated syndrome subjects

A coronal three-dimensional inversion recovery prepared FSPGR sequence was acquired with acquisition parameters TR=10.9ms, TE=4.2ms, TI=450ms, matrix 256x192, flip angle 20˚, FOV 240x180mm, resulting in 124 contiguous 1.5mm thick slices. Axial dual echo PD and T2-weighted images of the brain were acquired using a FSE sequence with acquisition parameters TR=3200ms, TE=19/95ms, matrix 256x256, flip angle 90˚, FOV 240x180mm, resulting in 46 contiguous 3mm thick slices. Gd-DTPA (Magnevist (Schering AG, Berlin, Germany)) (0.1mmol/kg) was injected, and approximately 15 minutes later T1-weighted images were acquired using a CSE sequence with acquisition parameters TR=600ms, TE=17ms, matrix 256x256, flip angle 90˚, FOV 240x180mm, resulting in 46 contiguous 3mm thick slices.

3.3.2 Relapsing remitting multiple sclerosis and control subjects

An axial three-dimensional inversion recovery prepared FSPGR sequence was acquired with acquisition parameters TR=10.9ms, TE=4.2ms, TI=450ms, matrix 256x160, flip angle 20˚, FOV 300x230mm, resulting in 124 contiguous 1.5mm thick slices. In addition a coronal three-dimensional inversion recovery prepared FSPGR sequence was acquired with acquisition parameters TR=10.9ms, TE=4.2ms, TI=450ms, matrix 256x192, flip angle 20˚, FOV 240x180mm, resulting in 124 contiguous 1.5mm thick slices. Axial dual echo PD and T2-weighted images of the brain were acquired using a FSE sequence with acquisition parameters TR=2000ms, TE=17/102ms, matrix 256x256, flip angle 90˚, FOV 240x180mm, resulting in 28 contiguous 5mm thick slices. T1-weighted images using a

CSE sequence were acquired with acquisition parameters TR=550ms, TE=20ms, matrix 256x256, flip angle 90˚, FOV 240x240mm, resulting in 28 contiguous 5mm thick slices. In patients, triple dose Gd-DTPA (0.3mmol/kg) was injected, and approximately 15 minutes later a further conventional SE sequence was acquired. None of the patients had experienced a clinical relapse or received a course of corticosteroids within a month prior to imaging.

3.3.3 Control subjects with same-day imaging

Two coronal three-dimensional inversion recovery prepared FSPGR sequences were acquired with acquisition parameters TR=15ms, TE=5.4ms, TI=650ms, matrix 256x256, flip angle 15˚, FOV 240x240mm, resulting in 124 contiguous 1.5mm thick slices.

3.4 Image analysis algorithms and methods

3.4.1 MIDAS

MIDAS software (Freeborough *et al*., 1997) is implemented in the C programming language running on a Unix platform. MIDAS allows three dimensional MR images to be displayed in simultaneous multiplanar views. Outlining of regions of interest (ROI) can be performed using both automated and semi-automated functions. The simultaneous display of orthogonal views allows the operator to edit the ROI in one view whilst it is updated in real time in the two other views, thereby aiding decisions regarding brain boundaries. Rigid-body registrations using Automated Image Registration (AIR) (Woods *et al*., 1998) can be implemented within the MIDAS software, and MIDAS provides an interface for quantification of brain atrophy using the BBSI.

3.4.2 FMRIB's Software Library (FSL)

The FMRIB's (Functional Magnetic Resonance Imaging of the Brain, www.fmrib.ox.ac.uk) software library includes image analysis and statistical tools for the analysis of both structural and functional MRI and is run on a Unix platform (Smith *et al*., 2004). Version 3.1 was used for all analysis. Structural image analysis tools used in this work include BET, an automated algorithm that segments brain from non-brain and models skull and scalp surfaces (Smith, 2002); input parameters can be altered to optimise segmentation. FMRIB's Automated Segmentation Tool (FAST) can be used for brain segmentation into different tissue types and bias field correction (Zhang *et al*., 2001a). FMRIB's Linear Image Registration Tool (FLIRT) provides linear inter- and intra-modal

registration (Jenkinson & Smith, 2001). SIENA is an automated algorithm for assessing structural brain change and estimating atrophy (Smith *et al*., 2002).

3.4.3 Statistical parametric mapping

SPM was developed by the functional imaging laboratory at the Institute of Neurology, London (Frackowiak *et al*., 1997). SPM99 was used running on a Matlab platform (MathWorks, Natick, Mass., USA). SPM software was applied to MR data to automatically delineate GM, WM and CSF.

3.4.4 DispImage

Dispimage, a display and image analysis package (D. Plummer, University College London Hospitals NHS Trust, UK) (Plummer, 1992) was run on a Unix system. The software allows the display of MRI in one plane and outlining of regions of interest using automated or semi-automated methods. Lesions on MR images were outlined using either an automated thresholding tool within Dispimage, which is based on intensity changes at the edge of the lesion, or manually, using a mouse-driven cursor.

3.4.5 Excalp

Excalp is an automated programme implemented in the C programming language running on a Unix platform (Yoo Done-Sik, University College London, UK. “Imaging and Segmentation of Bone in Neurological Magnetic Resonance”, PhD Thesis, 1998). This software was originally designed to perform an accurate segmentation of the skull from MR images, but in part achieves this through detection of the brain, and it has therefore been used as a brain segmentation tool. It is applied to individual image slices to strip the skull from brain images, by combining a histogram-based thresholding method and an edge detection method with connected components analysis and morphological operators.

3.5 Statistical analysis

STATA versions 8 and 9 (Stata Corporation, College Station, Texas, 1999) were used for statistical analysis of results.

4 Analysis and optimisation of brain atrophy measurement by the brain boundary shift integral in multiple sclerosis

4.1 Chapter introduction

With increasing focus on brain atrophy as a marker of disease progression in MS, and its potential for use as an outcome measure in clinical trials, automated and reliable measurement methods are currently attractive. Although the BBSI (Freeborough & Fox, 1997) is a validated technique, its application to patients with MS has been limited. The only study performed using this method showed increased brain atrophy rates in patients with RRMS, SPMS and PPMS relative to controls, but the number of subjects in each MS subgroup numbered less than 10 and disease duration ranged considerably between subjects (Fox *et al*., 2000b). In addition this method has not been applied to cohorts of subjects in the early stages of the disease who represent likely candidates for inclusion into future trials of treatments aimed at slowing the development of disability. Further investigation into the application and optimisation of the BBSI to larger cohorts of subjects early in the disease course is therefore warranted.

When designing studies investigating brain atrophy, the pulse sequence and parameters, and data processing protocol, need to be established and optimised in order to reduce the variability and increase the sensitivity of measurements. This may increase the power of the study to detect a difference in atrophy rates, for example between patients on treatment and placebo. This chapter investigates ways in which MR images and the BBSI method can be optimised for measurement of brain atrophy in patients presenting with a CIS and patients with early RRMS. Firstly the effect of image intensity normalisation, both within a single image and between serial images, is investigated. Secondly, optimal parameters specifying the location and width of the boundary sampling window in these cohorts are determined, and the influence of this on atrophy measures and statistical power investigated. Finally, this chapter explores whether averaging same-day data (images or atrophy rates) can improve measurement precision by the BBSI.

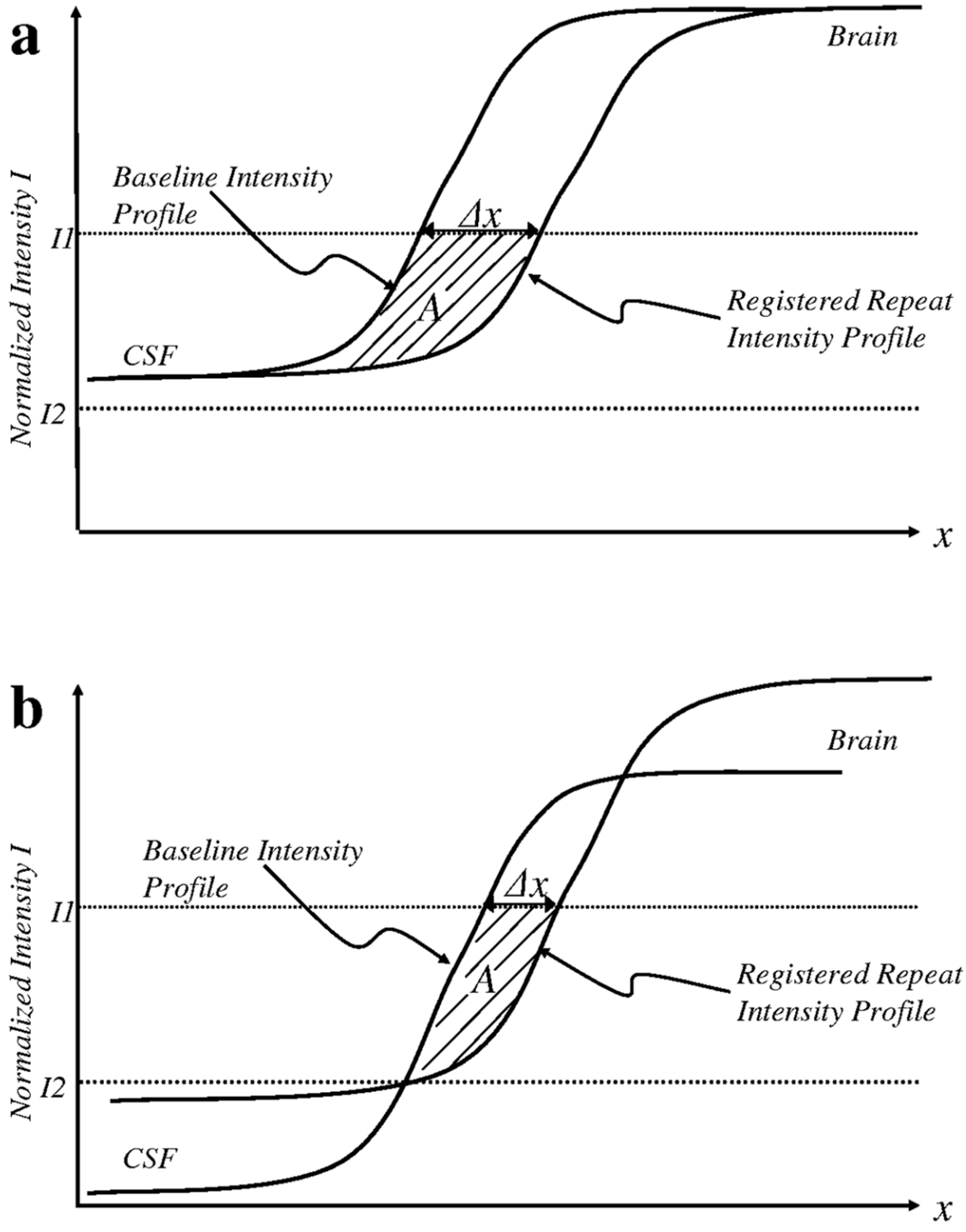
4.2 Assessment of non-uniformity correction using a) N3 and b) differential bias correction, on the brain boundary shift integral

4.2.1 Introduction

Underlying the BBSI is the assumption that any change in volume of the brain will be associated with an exactly equivalent shift in the brain boundaries. Following accurate registration of two brain images, areas of intensity loss between serial images are likely to represent shifts in tissue near the boundaries as brain atrophy occurs, and the BBSI derives an automated measurement of global atrophy by directly comparing the image intensity profiles of registered MR acquisitions. It requires selection of an intensity window which should be contained within all the intensity transitions associated with the boundaries of the brain (Figure 1-16), and this window should be standardised for a given scan type and cohort. As such, atrophy quantification by the BBSI could be invalid if images are affected by either intensity inhomogeneity (bias field, as described in Chapter 1.5.3) or differential intensity between serial images (differential bias).

A bias field will lead to a shift in the intensity profile of an image so that the intensity window may not be encompassed by the relevant boundary intensity transitions at only points on the brain boundary surface, i.e. the intensity profile may only partially sit within the given intensity window for some region(s), and therefore will contribute less to the estimate of atrophy. Figure 4-1a shows that intensity inhomogeneity in both baseline and repeat images has shifted the intensity profiles upwards. Consequently brain atrophy is underestimated as the boundary region is not entirely encompassed by the specified intensity window. This intensity profile represents just one point on the brain boundary and will therefore vary at different places in the boundary region of a single subject. In the presence of differential bias the intensity profile of one image is moved relative to the other, and similarly the intensity window may not encompass the boundary movement. In addition it is possible that non-boundary tissue could contribute to the measurement (Figure 4-1b). In addition to these potential effects of intensity bias, the BBSI requires segmentation of the brain both for image registration and for estimation of the brain boundary region. Typically this is done using MIDAS which requires specification of intensity thresholds to delineate the whole brain (Freeborough *et al*., 1997), and assumes intensity homogeneity within a tissue class. Furthermore, a bias field may make it difficult to recognise tissue boundaries for either setting of the thresholds or manual editing of regions.

Figure 4-1 *The effect of intensity inhomogeneity, on single scans and between serial images, on the BBSI. Based on a one-dimensional representation of the intensity profile through a brain boundary on serial imaging a) intensity inhomogeneity in both images has shifted the intensity profiles upwards and consequently the boundary shift (Δx) is underestimated as the boundary region is not entirely encompassed by the specified intensity window, b) differential bias has shifted the repeat image intensity profile relative to the baseline and consequently the boundary shift (Δx) is underestimated.*



As previously described, a number of methods have been developed to retrospectively correct for image intensity inhomogeneity (Chapter 1.5.3). One such method that is freely available on the Internet, is nonparametric nonuniform intensity normalisation (N3) (http://www.bic.mni.mcgill.ca/software/N3) (Sled *et al*., 1998). Validated on volumetric brain images acquired on a 1.5T scanner, it is a fully automated algorithm that estimates the distribution of voxel intensities within an image and then iteratively sharpens the resulting intensity histogram until the CV in the ratio between subsequent estimates of the nonuniformity field is below a given value. Although there are a number of other methods that could be used for the same purpose, N3 has been shown to be superior (Arnold *et al*., 2001). In addition, a study investigating a variety of processing streams on different acquisitions found that the choice of algorithm for reducing intensity bias had very little effect on tissue segmentation (Clark *et al*., 2006). Also N3 is a non-parametric method which may be important when attempting to correct for intensity inhomogeneity in patients with MS. WM lesions may cause greater intensity variations, but N3 can deal with this feature as this there is no limit to the number of tissue classes in the histogram.

Differential bias correction (DBC) has been described for the standardisation of intensity between two images from the same individual to control for differences in the bias field that may occur with serial imaging (Lewis & Fox, 2004). Calculation of the differential bias field is made on the assumption that in a difference image of registered serial brain images, the difference due to atrophy and noise is small-scale. A differential bias field is of a much larger scale and can be estimated by applying an appropriate filter to the difference image, and subsequently can be corrected for.

The aim of this study was to investigate the effect of N3 correction on individual images and DBC on serial images, and to determine their influence on the sensitivity and precision of brain atrophy measurements by the BBSI in patients with MS pathology and controls.

4.2.2 Methods

Subjects and imaging

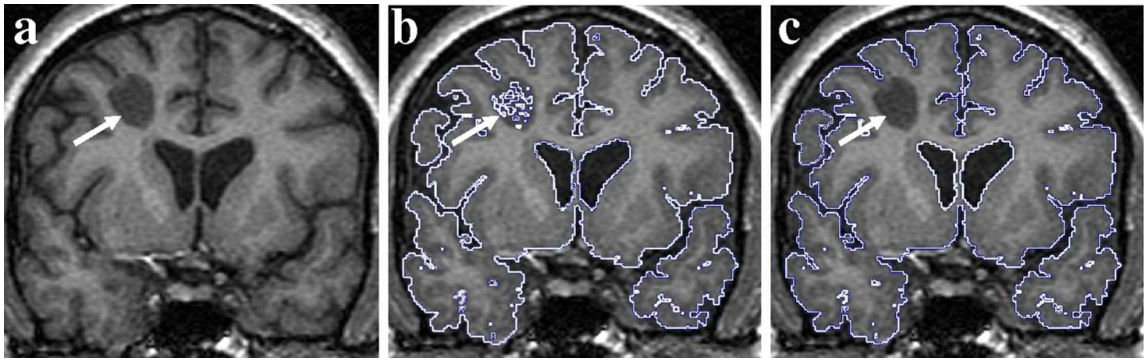
Eighty-four subjects with serial MR brain imaging ranging in age from 21-56 years were identified; 16 control subjects (seven male, mean age 35.1 years (SD 6.3)), 37 subjects who presented with a CIS (14 male, 35.0 years (SD 6.2)) and 30 subjects with RRMS

(eight male, 37.5 years (SD 7.4)). Subjects with RRMS were within four years of disease onset and had a median EDSS of 1.5 at baseline (range 0-3). All subjects had coronal FSPGR MRI at baseline and approximately one year later (mean follow-up time 1.14 years (SD 0.20)) according to the standard acquisition protocols described in Chapter 3.3.

MRI analysis

MIDAS was used to segment the brain on each image (Freeborough *et al*., 1997). Delineation of brain/non-brain was achieved by setting intensity thresholds to eliminate voxels that were greater or less than the given intensities, followed by erosions and dilations. Manual editing was performed when required. As WM lesions are often visible on T1-weighted MRI, appearing hypointense in relation to WM, whole brain segmentation using MIDAS may exclude these regions based on intensity thresholding (Figure 4-2). Subsequently, these regions may be classified erroneously as a brain boundary by the BBSI. Therefore these regions were filled in by manual editing of the ROI (Figure 4-2). All segmentations were performed blinded to subject identity and image time-point.

Figure 4-2 *The effect of hypointense lesions on whole brain segmentation by MIDAS. a) T1-weighted image showing lesion (arrow), b) the lesion is excluded from the brain ROI by MIDAS, c) following manual correction, the area of the lesion in included in the ROI.*



N3 was applied to each image in order to correct for intensity inhomogeneity within the images. Brain regions were then copied and pasted from the original images to the N3-corrected images and edited if necessary. To calculate the BBSI a registration algorithm (AIR) (Woods *et al*., 1998), determining the rotations, translations and shear required to obtain a subvoxel accuracy over the whole brain, positionally matched baseline and repeat

brains, and a linear scaling was used to account for variations in voxel size due to scanner drift. Spatial scaling factors were estimated from the brain to brain registration. The BBSI algorithm was applied to both uncorrected and N3-corrected image pairs. Each registration and subsequent atrophy calculation was performed with and without automated DBC, in addition to the normal intensity scaling that the BBSI carries out. Thus, four atrophy measures were obtained for each subject: i) no correction, ii) N3 correction only, iii) DBC only, iv) N3 correction and DBC.

Evaluation of the effect of N3 correction and DBC

Visual assessment of each registration was performed to check that it was accurate. In addition to analysing the influence on atrophy measures, the effect of N3 correction and DBC on the actual images was assessed. Visual assessment of images before and after each procedure was performed. The CV of WM intensity was determined in each image before and after N3 correction and compared in each subject group using a paired *t*-test. This required delineation of the WM on each baseline, non-corrected image using the brain region obtained from MIDAS, and FAST (Zhang *et al*., 2001a). This WM region was eroded once to avoid the inclusion of partial volume voxels at the GM/WM border. The baseline scan was registered (affine) to the repeat scan, and using the transformation parameters the eroded baseline WM mask was resliced to the repeat image space. The same baseline and repeat WM regions were applied respectively to baseline and repeat N3-corrected images (non-DBC corrected), and the mean and SD of voxel intensities within the WM region were calculated for each image. To assess the influence of DBC baseline brain voxel intensities were determined, and subsequently correlated with registered repeat brain voxel intensities (based on the MIDAS brain region). The correlation coefficient was determined before and after DBC and in each subject group a paired *t*-test was used to look for significant differences in these values.

Statistical analysis

Atrophy measures were corrected for scan interval and expressed as a percentage of baseline brain volume. The mean (SD) atrophy rates in each subject group before and after N3 correction and DBC were calculated. Mean atrophy rates obtained using the four different processing procedures were compared by fitting a model relating atrophy rate to procedure utilising a generalised estimating equations approach. The model assumed an exchangeable correlation structure and robust standard errors were calculated. This

approach allows for correlation between the measures and differences in the variance of measures between methods. A joint Wald test was used as an overall test of differences in mean atrophy rates between procedures. Where a significant difference was observed ($p<0.05$), paired *t*-tests were used to investigate pairwise differences. Likewise, a generalisation of Pitman's test to more than two observations was performed within each subject group to determine if there was a significant difference in the variance of atrophy rates when N3 correction and DBC were applied (Han, 1969). Where a significant difference was observed ($p<0.05$), Pitman's test was applied pairwise to observations.

The statistical power of any future trial using brain atrophy as an outcome measure will be driven by the mean and variance of measures in patients. For any two methods where a significant difference in the mean or variance of atrophy rates was found in the RRMS patient group, investigation into the relative statistical power was performed. The CV (σ/μ) of the two methods was calculated, as the square of this value is proportional to the sample size required for a randomised controlled trial with the power to detect a particular proportional reduction in brain atrophy rate. Hence the square of the ratio of the respective CVs indicates the relative number of patients required to detect a treatment effect using atrophy rates obtained using the different processing methods (for any expected percentage treatment effect and any required statistical power). A 95% bootstrap CI (bias corrected on the logarithmic scale) for the ratio was calculated using 1000 replicates. This CI was used to determine whether differences were statistically significant.

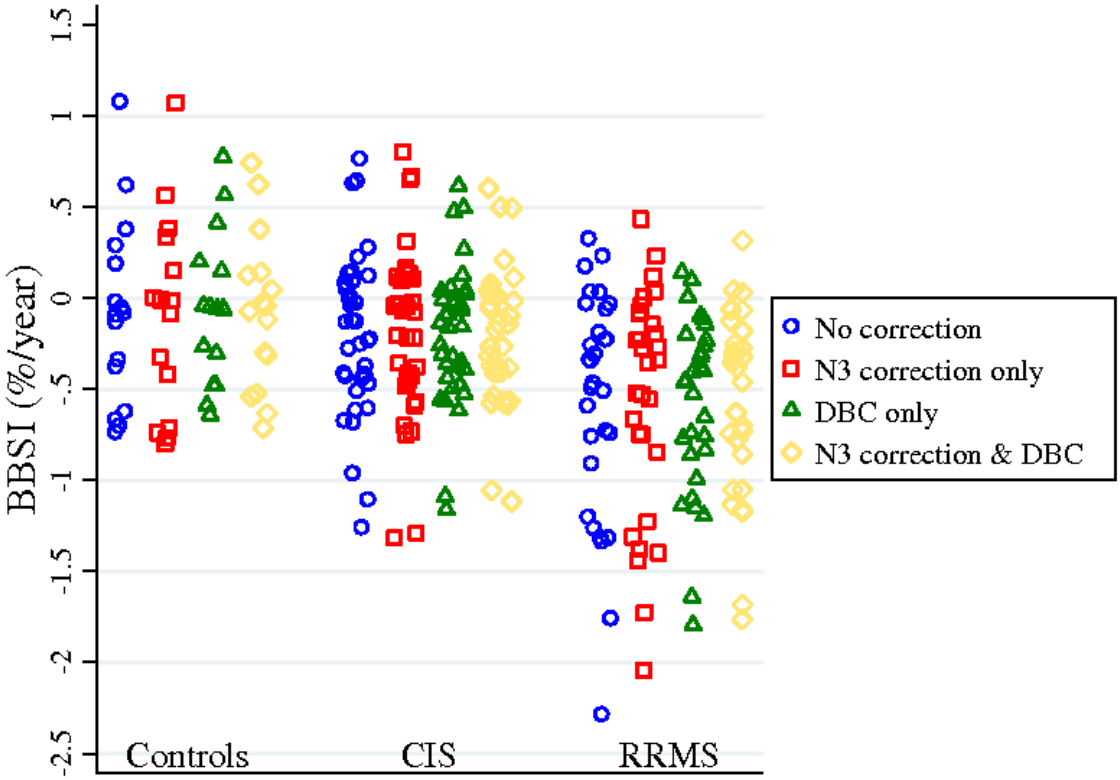
4.2.3 Results

On visual inspection none of the images had noticeable intensity inhomogeneity or differential bias, and following N3 correction it was not felt necessary to edit any of the brain regions. However comparing the CV of the WM in each image before and after N3 correction, it was found to be marginally but significantly reduced following N3 correction (mean difference -0.01120, SD 0.00654, 95% CI -0.01020 to -0.01221, $p<0.001$). In addition the correlation of voxel intensities in baseline and registered repeat brain regions was marginally but significantly greater following DBC than before (mean increase 0.00117, SD 0.00157, 95% CI 0.00151 to 0.00082, $p<0.001$). Mean (SD) BBSI atrophy rates for each subject group, with and without N3 and DBC correction are shown in Table 4-1 and Figure 4-3.

Table 4-1 *BBSI mean brain atrophy rates when images are uncorrected, or intensity inhomogeneity corrected using N3 and DBC, in controls and subjects with CIS and RRMS.*

		No DBC *Mean (SD) (% year^{-1})*	***DBC*** *Mean (SD) (% year^{-1})*
No N3 correction	*Controls*	-0.08 (0.51)	-0.07 (0.41)
	CIS	-0.18 (0.45)	-0.18 (0.38)
	RRMS	-0.56 (0.62)	-0.59 (0.49)
N3 correction	*Controls*	-0.08 (0.52)	-0.07 (0.42)
	CIS	-0.18 (0.46)	-0.19 (0.37)
	RRMS	-0.56 (0.62)	-0.59 (0.50)

Figure 4-3 *BBSI brain atrophy rates when images are uncorrected, or intensity inhomogeneity corrected using N3 and DBC, in controls and subjects with CIS and RRMS.*



Comparison of mean atrophy rates from all four processing methods demonstrated no significant differences in mean values when N3 correction or DBC were applied to images (controls p=0.86, CIS p=0.42, RRMS p=0.74). Applying the generalisation of Pitman's test to results from all four methods revealed a significant difference in the variance of atrophy rates ($p<0.001$ in each subject group). Pairwise Pitman's tests revealed that the variance of atrophy rates was significantly reduced in each subject group when DBC was applied to uncorrected or N3 corrected images (all $p<0.001$). No significant differences were observed when N3 correction was applied (p-values ranged from 0.16 to 0.66). Relative sample sizes to detect a particular proportionate difference were estimated to be 26% smaller using DBC (95% CI 37% to 18% reduction) compared with no correction, indicating significantly greater statistical power by application of DBC.

4.2.4 Discussion

In this study the effect of a nonparametric nonuniform intensity normalisation (N3) on single images, and differential bias correction (DBC) between serial images, on quantification of atrophy by the BBSI has been investigated. In the cohort studied it was confirmed that N3 correction significantly reduces the variability of voxel intensities within the WM, but correction did not significantly affect BBSI-derived atrophy measures. However DBC significantly increased correlation between brain voxel intensities on serial images, and led to a significant decrease in the variance of atrophy rates quantified by the BBSI, which resulted in greater statistical power.

In a previous study that applied the BBSI to MS subjects, N3 correction and DBC were not applied (Fox *et al*., 2000b). On visual inspection of the images included in the present study intensity bias was not evident, yet a significant decrease in the variability of WM intensity was detected following application of N3 correction. Although the true amount of nonuniformity in any MRI image is unknown and it is not possible to measure the accuracy of N3 correction directly, this suggests that inhomogeneity correction should be applied routinely to images. However, it was shown that N3 correction did not significantly affect either the mean or variance of atrophy measures suggesting that the BBSI is robust to small amounts of intensity bias. This finding is similar to that observed in a study applying a boundary shift integral method in controls and patients with Alzheimer's disease; N3 bias correction had little effect on brain atrophy quantification

and did not influence group separation (Gunter *et al*., 2003). However bias correction might be important if investigating serial changes in WM volumes, and this warrants further investigation.

Previously it has been shown that the BBSI can be considerably altered when there is a change in contrast between serial images (Preboske *et al*., 2006), and efforts should be made to standardise the acquisition over time. Again, visual assessment of images did not highlight any differences in intensity between any of the image pairs, but the correlation between the intensity of voxels on serial images increased following DBC. Although correlation does not necessarily mean that the intensity values agreed, only that they were more strongly associated, analysis of atrophy rates showed that DBC significantly decreased the variance of measures and consequently increased the statistical power. That DBC influenced atrophy measures to a greater extent than N3 correction, is likely to have been because the BBSI depends to a greater extent on the difference in voxel intensity between serial images than the consistency of intensity of a tissue class. Indeed if both images have a similar bias field (Figure 4-1a) there is likely only to be an under-estimation of atrophy. However with differential bias, noise will be added to the measurement that will increase measurement variability. Unlike the original paper describing DBC (Lewis & Fox, 2004), there was no apparent decrease in the mean atrophy rate following the application of DBC, which was suggested to be due to some atrophy being smoothed away during the processing. This may not have been so apparent in the present study due to the lower atrophy rates observed compared with the original study which was performed in patients with established Alzheimer's disease.

If both N3 correction and DBC are to be applied prior to the BBSI, one important consideration may be the order in which they are applied and the influence of one on the other. One may assume that it is logical to improve intensity homogeneity on individual scans prior to intensity standardisation. Indeed a study which investigated this aspect of optimisation of images found that this was the best order to perform procedures (Madabhushi & Udupa, 2005).

One of the limitations of this study is that the influence of N3 correction on the initial brain segmentation and subsequent atrophy quantification was not investigated, as the brain ROI was copied from an uncorrected image to the N3-corrected image. This aspect

would have been difficult to examine however, given that there is a degree of intra-rater error associated with repeated brain segmentations, making it difficult to determine differences due to measurement error and those due to differences in boundary recognition. Moreover, a study that investigated the effect of intensity bias correction on brain segmentation from four different automated and semi-automated algorithms found that it did not affect method performance (Fennema-Notestine *et al*., 2006), and is therefore unlikely to have influenced the whole brain segmentations in this study. It should also be noted that although N3 should be insensitive to pathology such as lesions, this study did not investigate whether there is a threshold beyond which the normalisation process may be significantly influenced by hypointense WM lesions.

This study demonstrates that a) N3 correction significantly increases intensity homogeneity on individual images and b) DBC significantly increases intensity homogeneity between serial images. Most importantly, DBC appears to decrease the variance of atrophy rates obtained using the BBSI suggesting increased measurement precision. The relative reduction in sample size of one quarter that was a consequence of this would greatly facilitate the performance of exploratory trials using brain atrophy as the outcome measure. Although N3 correction did not influence atrophy rates it is likely that there is a threshold of inhomogeneity beyond which N3 correction could be a valuable pre-processing step for the BBSI. As N3 correction and DBC are fully automated methods that require minimal operator time, these findings suggest that these processes should be performed routinely in studies prior to atrophy quantification by the BBSI.

4.3 Selection of optimal parameters for quantification of the brain boundary shift integral

4.3.1 Introduction

When calculating the BBSI, the range of intensities (I_1, I_2) over which the integral is calculated must be defined. This intensity range or window has a centre $I_c = (I_1 + I_2)/2$, and width $I_w = I_1 - I_2$ (Freeborough & Fox, 1997). These parameters are user defined and the intensity window should maximise the number of contributing boundary voxels whilst falling entirely within all of the intensity transitions associated with the boundaries between brain and CSF. These opposing requirements are complicated by the fact that the brain is made up of tissues with differing intensities. As shown in Figure 1-16 the boundary shift is estimated as the area A between two intensity profiles within the defined

window, divided by the intensity window I_w. A window that is too wide is likely to underestimate brain atrophy as the intensity window falls outside the brain boundary and therefore the normalising factor (I_1-I_2) is disproportionately large relative to A. By using a narrower intensity window there will be a greater chance that it falls entirely within the intensity transitions of the brain boundary and atrophy will be more reliably detected. However too small a window can lead to the window lying entirely outside the intensity transitions too often, resulting in measurement error.

The optimum window settings are primarily dependent on the scan acquisition protocol which can differ between centres and study cohorts. The arrangement and intensity of adjacent brain tissues will not vary significantly between subjects from one cohort or scanner. The optimum window settings can therefore be determined for a given scanner and scan type and then applied to multiple subjects. The default parameters of I_1=0.75 and I_2=0.25 were those initially published (where 0.75 refers to 75% of mean normalised brain intensity) (Freeborough & Fox, 1997). Optimised parameters (for the particular image acquisition) can be estimated by comparing the BBSI to segmented brain volume differences (BVD), for a range of window parameters, and maximising their association.

The objective of this study was therefore to determine the optimal window settings for maximising the sensitivity, precision and accuracy of the BBSI on i) coronal and ii) axial T1-weighted volumetric acquisitions, in the cohort of patients and controls being investigated in this thesis.

4.3.2 Methods

Subjects and imaging

Eighty-eight subjects were identified with coronal FSPGR MRI at baseline and approximately 12 month follow-up (mean follow-up time 1.16 years (SD 0.24)); 16 control subjects (seven male, mean age 35.1 years (SD 6.3)), 42 subjects presenting with a CIS (14 male, 34.4 years (SD 6.3)) and 30 patients with RRMS (eight male, 37.5 years (SD 7.4)). In addition, 17 control subjects (seven male, mean age 35.2 years (SD 6.2)) and 29 subjects with RRMS (eight male, mean age 37.7 years (SD (7.5)) had axial FSPGR MRI at baseline and approximately 12 month follow-up (mean 1.06 years (SD 0.11)). Imaging was performed according to the standard acquisition protocols described in Chapter 3.3.

MRI analysis

The brain was segmented on each image using MIDAS as described in Chapter 4.2.2, and N3 correction was applied. Brain volumes were determined from the segmentations and the difference in volume between the baseline and repeat image was calculated. The BBSI algorithm, with DBC and default parameters (I_c=0.5 and I_w=0.5, i.e. I_1=0.75 and I_2=0.25), was applied to each serial image pair. Normalisation of signal intensities within the images was achieved by dividing the intensity of each voxel by the mean intensity of the internal intersection region (the intersection between the baseline and registered repeat segmented brain regions, eroded once). This approximately scales the voxel intensities within the boundary region between zero and one. Visual assessment of each registration was performed to check registration accuracy.

Subsequently, atrophy was estimated from all registered image pairs using different window settings. An automated algorithm was designed for this purpose and required only specification of I_1 and I_2. It was hypothesised that reducing the window width (I_w) would increase the sensitivity of the BBSI. Therefore based on a window width of 0.3, BBSIs were calculated for window centres (I_c) ranging from 0.4 to 0.8 (at 0.05 intervals). Segmented volume differences and atrophy measures derived from the BBSI were corrected for scan interval and expressed as a percentage of baseline brain volume.

Statistical analysis

Data from coronal and axial images were analysed separately. For each window setting atrophy rates calculated using the BBSI were plotted against the difference in segmented brain volume obtained from MIDAS, and linear regression lines were fitted to the data (the values should approximate each other). In each subject group mean (SD) atrophy rates were calculated for each window setting tested. In patients with RRMS effect sizes (μ/σ) were calculated in order to determine the statistical power of atrophy measures from each window setting.

4.3.3 Results

The linear regression lines fitted to data from coronal images, comparing the segmented volume difference and BBSI measures obtained using different window settings, are shown in Figure 4-4. Using a window width of 0.3 it was found that a window centre of 0.55 gave the closest approximation to the segmented volume difference. The window

width was reduced to 0.2 at this window centre which increased the agreement of the BBSI and segmented volume difference further. These results can be demonstrated by looking at a real example of an intensity profile, seen in Figure 4-5, which shows the boundary transitions that fall within different intensity windows.

Figure 4-4 *Correlation of segmented volume difference and BBSI brain atrophy rates, obtained using different window settings on coronal T1-weighted volumetric images. Atrophy rates from controls and subjects with CIS and RRMS were plotted and regression lines fitted to the data. Both the window centre (c) and the window width (w) lie between 0 and 1.*

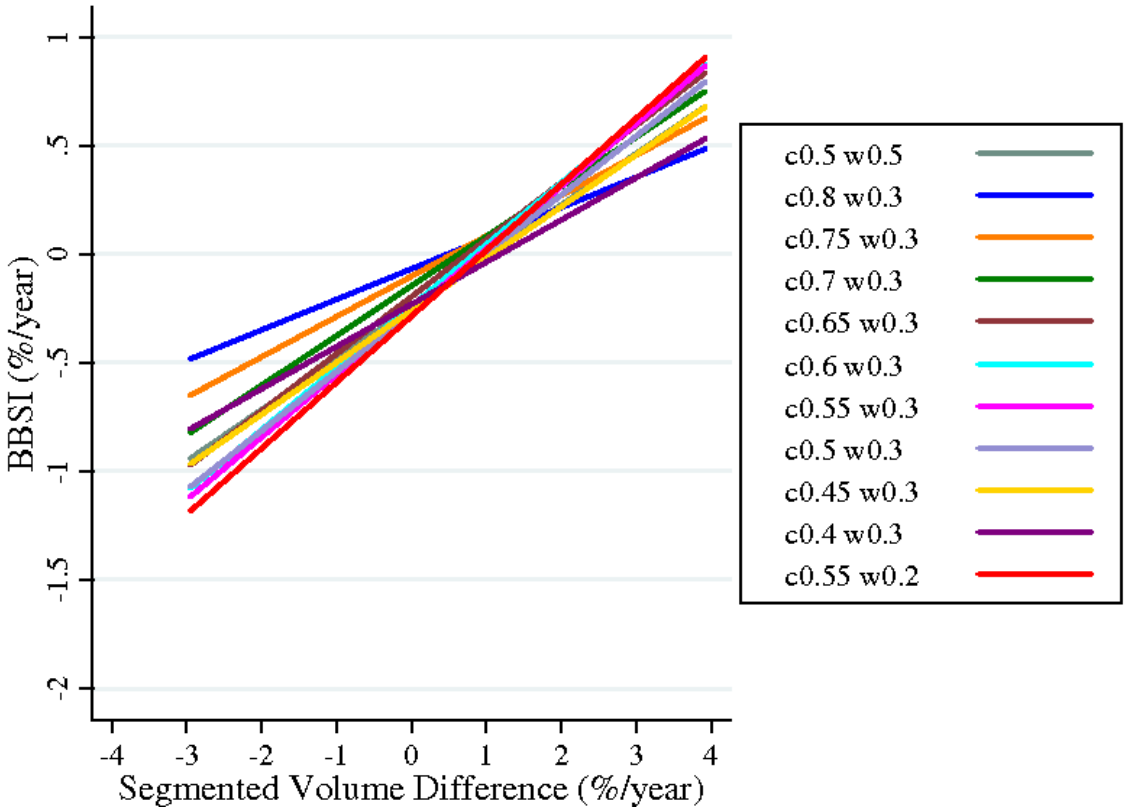


Figure 4-5 *The effect of different intensity window settings on BBSI atrophy measurement. Different BBSI intensity windows are overlaid on a one-dimensional intensity profile through a brain, showing how the window relates to the intensity transitions of the brain boundary. The intensity profile was generated by drawing a line one voxel high between the points A and B (top image), and normalising the intensity value of each voxel in this region to the mean whole brain intensity.*

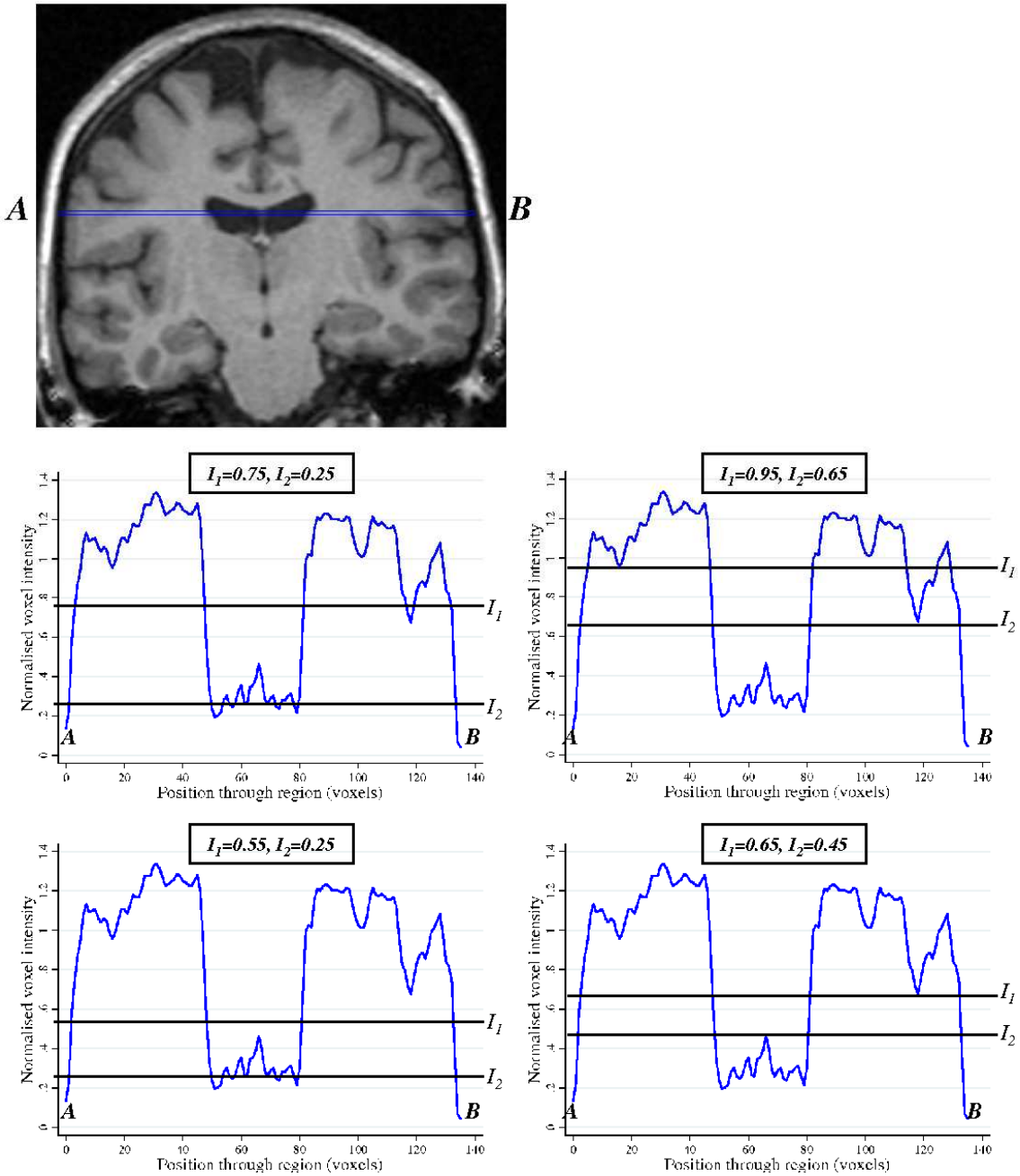


Table 4-2 shows the mean atrophy rates in each subject group for the different intensity window parameters tested. When looking at the different window centre settings for a window width of 0.3, mean atrophy rate peaked in patient groups when the window centre was set at 0.55. Mean atrophy rate increased when the window width was reduced to 0.2 at this window centre. Mean atrophy rate was closest to zero in control subjects for a window centre of 0.6. As mean atrophy rate increased, the variance of measures increased also. However the effect sizes for patients with RRMS reveal that these increases were not proportional, as the effect size was at its highest value when a window centre of 0.5 was utilised (Table 4-2).

Table 4-2 *BBSI mean brain atrophy rates quantified using different intensity window settings for coronal T1-weighted volumetric images, in controls and subjects with CIS and RRMS. Effect sizes (μ/σ) are given for subjects with RRMS.*

Window centre	***Window width***	***Mean (SD) atrophy rate (% year^{-1})***			***Effect size***
		Controls	***CIS***	***RRMS***	
0.5 (default)	0.5 (default)	-0.07 (0.42)	-0.18 (0.38)	-0.59 (0.50)	1.17
0.8	0.3	0.06 (0.26)	-0.03 (0.27)	-0.28 (0.32)	0.86
0.75	0.3	0.06 (0.33)	-0.06 (0.34)	-0.37 (0.40)	0.92
0.7	0.3	0.05 (0.39)	-0.09 (0.40)	-0.48 (0.48)	0.99
0.65	0.3	0.03 (0.44)	-0.13 (0.48)	-0.58 (0.54)	1.08
0.6	0.3	0.00 (0.48)	-0.17 (0.47)	-0.65 (0.57)	1.15
0.55	0.3	-0.03 (0.49)	-0.19 (0.47)	-0.68 (0.58)	1.18
0.5	0.3	-0.06 (0.48)	-0.19 (0.44)	-0.66 (0.56)	1.19
0.45	0.3	-0.08 (0.44)	-0.19 (0.39)	-0.61 (0.52)	1.18
0.4	0.3	-0.09 (0.38)	-0.16 (0.32)	-0.52 (0.45)	1.15
0.55	0.2	-0.04 (0.52)	-0.21 (0.49)	-0.72 (0.60)	1.21

The linear regression lines fitted to data obtained using axial images, comparing the segmented volume difference and BBSI measures obtained using different intensity window settings, is shown in Figure 4-6. As with coronal images, using a window width of 0.3 it was found that a window centre of 0.55 gave the closest approximation to the segmented volume difference for brain atrophy rate. Window width was reduced to 0.2 at

this window centre which increased the agreement between the BBSI and segmented volume difference still further, when atrophy was detected.

Figure 4-6 *Correlation of segmented volume difference and BBSI brain atrophy rates obtained using different window settings on axial T1-weighted volumetric images. Atrophy rates from controls and subjects with RRMS were plotted and regression lines fitted to the data. Both the window centre (c) and the window width (w) lie between 0 and 1.*

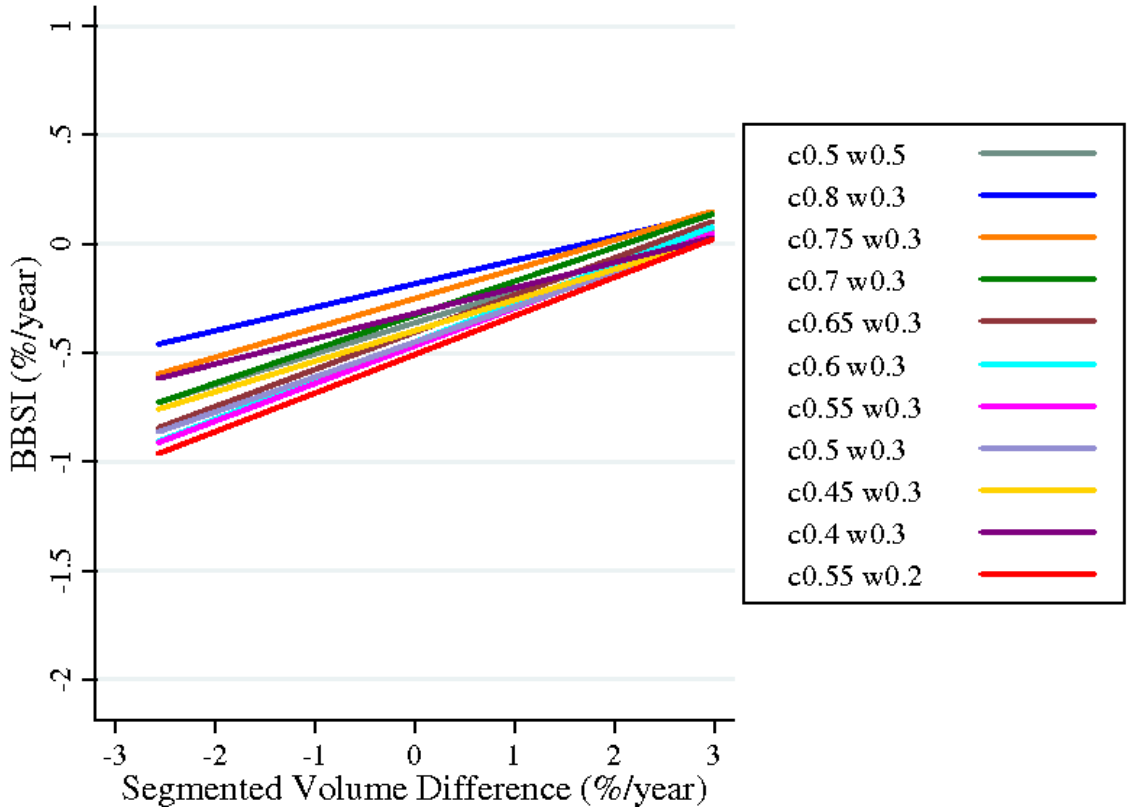


Table 4-3 shows the mean brain atrophy rates in controls and RRMS subjects for the different intensity window settings. When looking at the different window centre settings for a window width of 0.3, mean atrophy rate peaked in the RRMS group when the window centre was set at 0.55. When the window width was decreased to 0.2 at this window centre, mean atrophy rate increased. In controls mean atrophy rate was closest to zero for a window centre of 0.75. As with coronal images, as mean atrophy rate increased the variance of measures increased, but again this was not proportional as the effect sizes for patients with RRMS reveal that it was highest for a window centre of 0.55.

Table 4-3 *BBSI mean brain atrophy rates quantified using different intensity window settings for axial T1-weighted volumetric images, in controls and subjects with RRMS. Effect sizes (μ/σ) are given for subjects with RRMS.*

Window centre	***Window width***	***Mean (SD) atrophy rate (% year^{-1})***		***Effect size***
		Controls	***RRMS***	
0.5 (default)	0.5 (default)	-0.08 (0.37)	-0.54 (0.42)	1.29
0.8	0.3	0.02 (0.23)	-0.31 (0.33)	0.94
0.75	0.3	0.00 (0.29)	-0.41 (0.40)	1.03
0.7	0.3	-0.02 (0.35)	-0.52 (0.46)	1.15
0.65	0.3	-0.05 (0.40)	-0.63 (0.49)	1.28
0.6	0.3	-0.07 (0.44)	-0.69 (0.52)	1.33
0.55	0.3	-0.09 (0.46)	-0.71 (0.52)	1.37
0.5	0.3	-0.10 (0.45)	-0.67 (0.49)	1.37
0.45	0.3	-0.10 (0.41)	-0.58 (0.44)	1.32
0.4	0.3	-0.09 (0.36)	-0.46 (0.37)	1.24
0.55	0.2	-0.10 (0.48)	-0.76 (0.54)	1.42

4.3.4 Discussion

This study aimed to identify the optimal intensity window parameters with which to calculate the BBSI on both coronal and axial FSPGR acquisitions acquired in the cohort under study. It was found that an intensity window of I_1=0.65 and I_2=0.45 (i.e. I_c=0.55 and I_w=0.2) provided the most sensitive measure of brain atrophy. Although the variance of measures was increased at this value, this was proportionately smaller than the increase in sensitivity, which meant that greater statistical power could be achieved using these parameters, as measured by the effect size in patients.

The optimal parameters determined for both the coronal and axial images were the same, but this is not unexpected given that the scans were acquired on the same MR scanner using a similar protocol. Another study which looked at the optimum window settings for a dataset obtained on the same scanner as that used in the current study, similarly found that the best parameters were an intensity window between I_1=0.7 and I_2=0.5 (i.e. I_c=0.6 and I_w=0.2) (Boyes *et al*., 2004). The marginal difference in parameters between these studies may have been due to the fact that the current study included images from patients

with MS, and hypointense periventricular lesions may have influenced the results. However in addition to parameters being dependent on the scan dataset they may also vary for data acquired on different scanners. This study has shown how atrophy rates can vary considerably depending on the window parameters chosen. Although the optimal parameters did not influence effect size to a great extent relative to default parameters, changes were nonetheless observed, and could be more significant for other datasets. This suggests that it may be advantageous to optimise window parameters for any given dataset, particularly for clinical trials in which even small improvements in the sensitivity or precision of atrophy measurements could improve the ability to detect a treatment effect. Determining optimal parameters can be largely automated with minimal operator interaction required. As this was a retrospective investigation, all subjects with available imaging were included in the study to give power to the results and optimise over a wide range of atrophy rates. For a prospective study or clinical trial however, optimisation could be performed on a subset of subjects which would minimise the amount of computer processing time required to assess a number of different window settings on larger cohorts.

It should be noted that although increased power was gained from reducing the window width to 0.2, it was felt that the window should not be made any smaller. There is the potential that reducing I_w further could lead to spurious results in some instances, when the window could possibly fall entirely outside many brain boundary shift areas. With a wider window setting there is increased confidence that some of the brain boundary shift will lie within the window.

One of the limitations of this study is that not every possible window setting was tested, and it was not assessed whether atrophy rates for a window centre (I_c) of 0.1 would be realistic. For clinical trials, a wider range of intensity windows could be investigated, including narrower windows, and the results looked at carefully in correspondence with the difference images to determine whether results appear realistic. A further limitation of this investigation is that BBSI atrophy rates were assessed against the BVD. This measure may be subject to error itself and it is not possible to judge whether it represents the best estimate of overall brain atrophy. Indeed with axial acquisitions, positive atrophy rates or brain “growth” did not agree well between segmented volume difference and the BBSI using any of the window parameters. However even without approximating BBSI

measures to the volume difference, the parameters which give the most realistic, sensitive and precise measure should be identified. As such, an intensity window between I_1=0.65 and I_2=0.45 (i.e. I_c=0.55 and I_w=0.2) has been shown to increase the sensitivity and power of brain atrophy measurements by the BBSI compared with default parameters in both coronal and axial FSPGR images. For the remainder of this thesis these parameters will be applied in other investigations using the BBSI in this cohort.

4.4 Measurement of brain atrophy by the brain boundary shift integral on different volumetric acquisitions and average images

4.4.1 Introduction

With most atrophy measurement techniques, optimisation of MR images for the technique may increase the sensitivity and precision of atrophy measures. This is important when designing large prospective studies of brain atrophy, as the acquisition parameters and sequence for images can be decided upon prior to acquiring data in order to maximise the reliability of measures. Whilst the BBSI has been applied to numerous T1-weighted volumetric MRI, these images can vary in appearance depending on the MRI scanner used for acquisition and individual scan parameters (Fox *et al*., 2000b; Fox *et al*., 2005; Henley *et al*., 2006; Schott *et al*., 2005). Images that maximise SNR and CNR, and minimise chemical shift artefact, are likely to produce more reliable results. However no direct comparison of atrophy rates quantified on different acquisitions has been performed, so it is unclear to what extent this may effect measured atrophy rates. For multicentre studies, it is important to ensure that the BBSI is robust to small variations in images between subjects caused by differences in MRI hardware and acquisition parameters. In this study MR data has been acquired at one site only, so it is not possible to compare atrophy measures by the BBSI on acquisitions from different sites. However two FSPGR images have been acquired on the same day in controls and patients with RRMS. Although acquisition parameters for the two images were similar, the voxel dimensions and acquisition plane of the images are different. Therefore the first aim of this study was to investigate relative atrophy quantification by the BBSI on these two T1-weighted volumetric acquisitions and determine whether the BBSI was robust to these small differences and whether there was any systematic bias.

In addition to optimising data acquisition by altering scan parameters, averaging repeated data could increase measurement precision. Noise introduced by either the MRI hardware

or the atrophy measurement technique could be reduced through this process as the signal is effectively doubled. Same-day imaging allows repeated atrophy measurements to be obtained and results can subsequently be averaged. A slightly more complex approach involves averaging the images themselves prior to analysis, which should increase the SNR by a factor of 1.41. It has previously been shown that the increased signal that can result from averaging data at each spatial location may increase SNR (Holmes *et al*., 1998). This may allow more accurate and precise brain atrophy measurements to be made. Indeed one study investigating the impact of different acquisition protocols and processing streams on tissue segmentation used an average image as the gold standard (Clark *et al*., 2006). Averaging of volumetric images has also been utilised in Freesurfer, where a number of T1-weighted images can be combined prior to determination of cortical boundaries (Salat *et al*., 2004). The second objective of this study was therefore to determine whether averaging a) same-day data (BBSI measures) or b) same-day images, increased measurement precision of the BBSI. The former involves two separate BBSI measurements from two different scan pairs where two baseline and two repeat images were obtained on the same day. The latter involves merging images acquired at the same time-point and performing one BBSI measurement between the baseline average image and the repeat average image.

4.4.2 Methods

Subjects and MR imaging

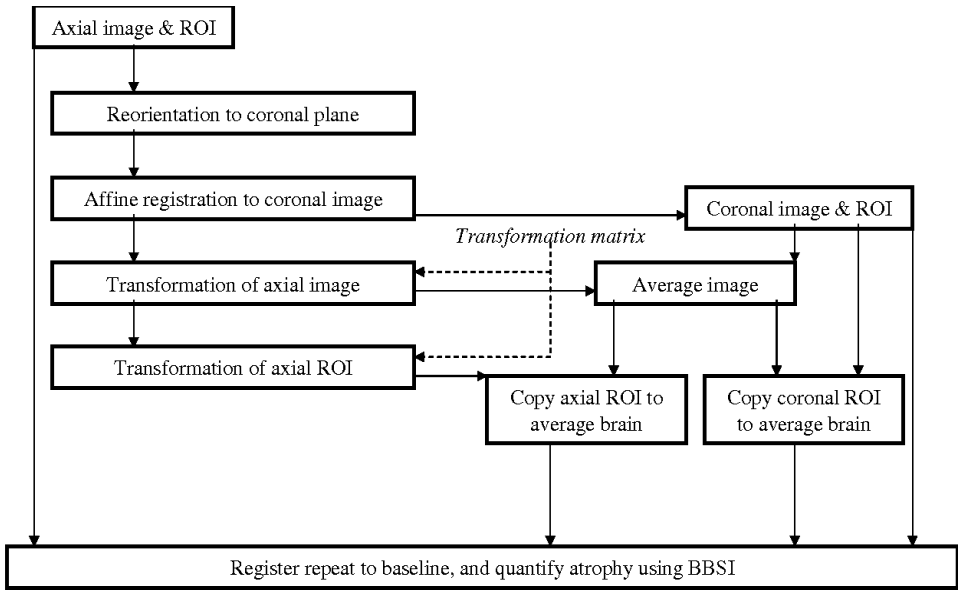
Twenty-nine subjects with RRMS (nine male, mean age 37.5 years (SD 7.5)) and 16 control subjects (seven male, mean age 35.1 years (SD 6.3)) were identified who had T1-weighted volumetric FSPGR images acquired in both the coronal and axial planes at baseline and approximately one year follow-up (mean 1.06 years (SD 0.11)). The imaging protocols described in Chapter 3.3.2 were used for MRI acquisition, and coronal and axial images were acquired in the same scan session. Voxel dimensions were 0.94 x 0.94 x 1.5mm^3 in the coronal image and 1.17 x 1.17 x 1.5mm^3 in the axial image.

MRI analysis

Whole brain segmentation was performed on all images using MIDAS, as described in Chapter 4.2.2, and N3 correction was applied. Affine registration was used to transform each repeat image to the baseline image (coronal to coronal and axial to axial) and the BBSI quantified using DBC and the optimised parameters determined in Chapter 4.3.

Figure 4-7 gives an overview of the procedure applied to average the axial and coronal image for a given subject and time-point. The axial image and region were first reoriented to the coronal plane whilst maintaining all voxel dimensions. The reoriented axial image was registered to the coronal image using affine registration, and re-sampled to give voxel sizes equivalent to the coronal image. An in-house algorithm was used to average the two co-registered images. Image intensity was normalised by dividing each image through by its mean intensity over the whole brain region, and then image intensities were rescaled between 0 and 1000. The transformation matrix used to register the axial image to the coronal image was applied to the axial brain ROI. This transformed axial brain ROI and the coronal brain ROI were separately copied to the average image. Each average repeat image was then registered to the corresponding average baseline image, using affine registration, and the BBSI was quantified. To avoid biasing results, registration and BBSI quantification was performed twice on the averaged images, once using the axial brain ROIs and once using the coronal brain ROIs. Again, DBC and the optimised parameters determined in Chapter 4.3 were applied. Visual assessment of all registrations was performed to check accuracy.

Figure 4-7 *Overview of the procedure used for creating an average image and calculation of the BBSI on average, axial and coronal images.*



Intensity profiles for the coronal, axial and average images of a randomly chosen subject were generated. Using MIDAS a line one voxel high was positioned across each image (which were registered to the same space), and the intensity at each voxel was

determined. The brain region, eroded once, was used to quantify mean brain intensity, and voxel intensities of the line were normalised to this value. In addition, the SNR was measured for each image. For each subject it was ensured that all images were in the same space as the coronal baseline image. Due to the procedures outlined and performed previously, only registration of the coronally reformatted repeat axial image to the reformatted and registered (to coronal) axial baseline image was required. On each coronal baseline image a region of approximately 1000 voxels was drawn within the WM of the anterior frontal lobe. It was ensured that this region did not include any GM or lesions. For coronal, axial and average images a difference image was generated from the registered baseline and repeat images for each subject. SNR was calculated according to Equation 4.1 (Price *et al*., 1990).

$$SNR = \frac{\sqrt{2}\ \text{mean signal in WM ROI of image}}{\text{mean signal in WM ROI of difference image}} \qquad (4.1)$$

Statistical analysis

SNR measurements from the different images (coronal, axial and average) were compared pairwise using a paired *t*-test. All BBSI values were corrected for scan interval and expressed as a percentage of baseline brain volume. A Bland-Altman plot of BBSI values obtained on coronal and axial acquisitions was generated to investigate their association. The BBSI quantified on serial coronal and axial images was averaged and subsequently mean (SD) atrophy rate in controls and RRMS subjects was determined for i) coronal images, ii) axial images, iii) average of coronal and axial atrophy rate, iv) average images using coronal brain regions, v) average images using axial brain region. As in Chapter 4.2 mean atrophy rates obtained on the three different images and average values were compared by fitting a model relating atrophy rate to method, utilising a generalised estimating equations approach. The model assumed an exchangeable correlation structure and robust standard errors were calculated. A joint Wald test was used as an overall test of differences in mean atrophy rates between methods. Where a significant difference was observed ($p<0.05$), paired *t*-tests were used to investigate pairwise differences. Again, a generalisation of Pitman's test to more than two observations was performed within each subject group to determine if there was a significant difference in the variance of atrophy rates using the different images and

methods (Han, 1969). Where a significant difference was observed ($p<0.05$), Pitman's test was applied pairwise to observations.

Investigation into the relative statistical power of any two methods, where a significant difference in the mean or variance of atrophy rates was found, was performed. The CV (σ/μ) of the two methods was calculated, and the square of their ratio determined to indicate the relative number of patients required to detect a treatment effect using the two different methods.

4.4.3 Results

Upon visual inspection, all the registrations appeared acceptable. An example of an original coronal and axial image and the average image generated is given in Figure 4-8. The average images appeared to be smoother with greater intensity homogeneity within each tissue class. The mean SNR was significantly greater in axial images than in coronal images (mean 38.4 versus 29.0, $p<0.001$) and significantly greater in average images (mean 46.1) than either coronal or axial images (both $p<0.001$).

Looking first at whether there was any systematic bias between atrophy rates obtained on the coronal and axial T1-weighted volumetric acquisitions, Figure 4-9 shows the Bland-Altman plot of the average of these values and the difference between them for each subject. The mean difference in atrophy rates between the two acquisitions (coronal – axial) was 0.03% $year^{-1}$ (SD 0.58, 95% CI -0.28 to 0.34, $p=0.85$) in controls and 0.03% $year^{-1}$ (SD 0.50, 95% CI -0.16 to 0.22, $p=0.73$) in patients with RRMS. Therefore no bias in average measures was evident, although differences up to approximately 1% $year^{-1}$ were observed within individual patients. The 95% reference range shows the values within which 95% of the differences between measures from coronal and axial images are expected to lie, and it is interesting to note that this range is wider in control subjects.

Figure 4-10 shows that the intensity profile of the average image lay between those of the coronal and axial images. Figure 4-11 shows the scatterplot of atrophy rates obtained for all subjects, whilst Table 4-4 gives the mean atrophy rates in each subject group when quantified using i) coronal images, ii) axial images, iii) average coronal and axial atrophy rate, iv) average image using coronal brain regions, v) average axial image using axial brain regions.

Figure 4-8 *Visual comparison of a coronal, axial and average image from a single subject. Images are shown when viewed in both the coronal and axial plane. From left to right: the original coronal image, the original axial image, the axial image following reorientation and registration to the coronal image, the average image.*

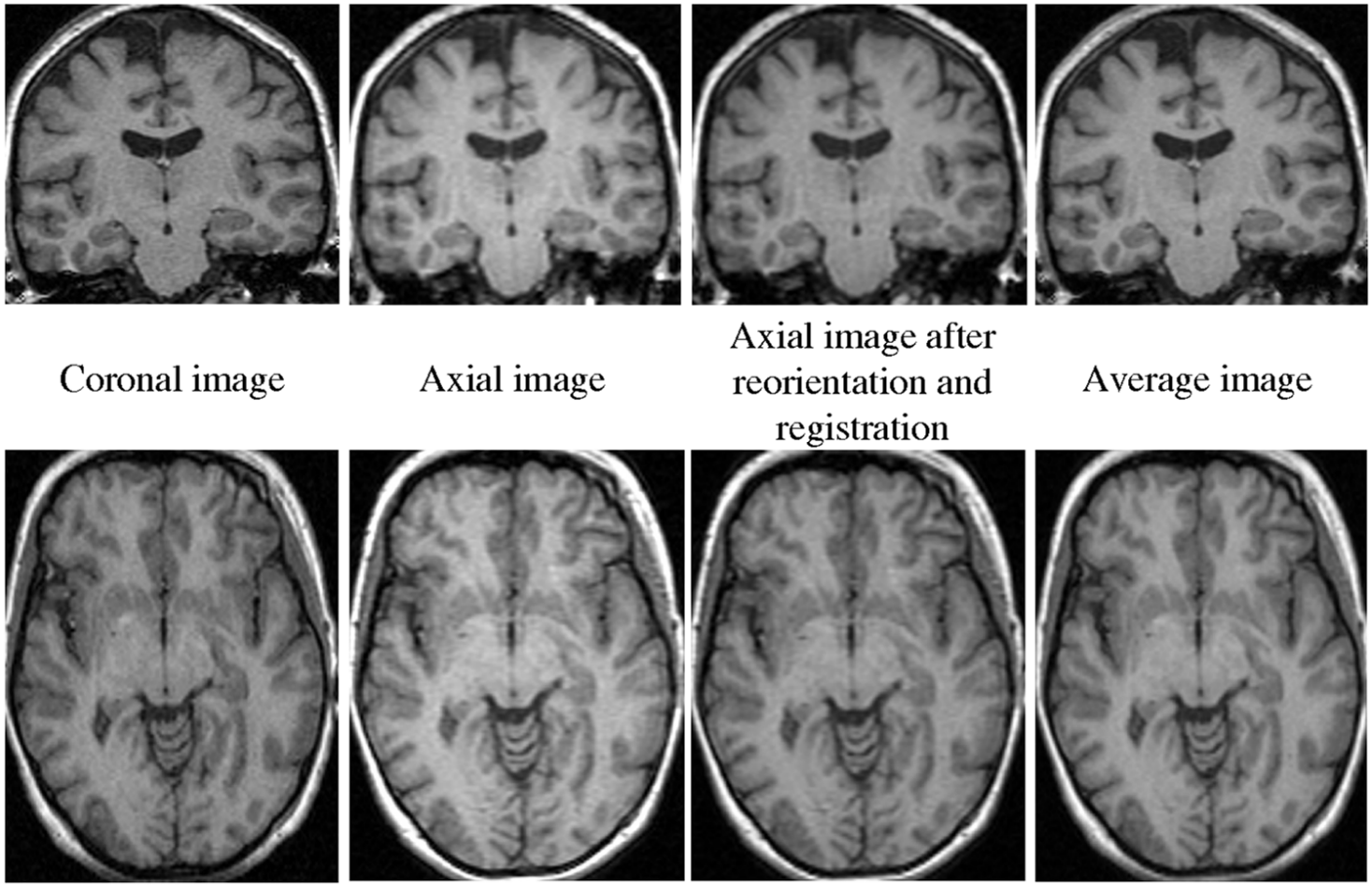
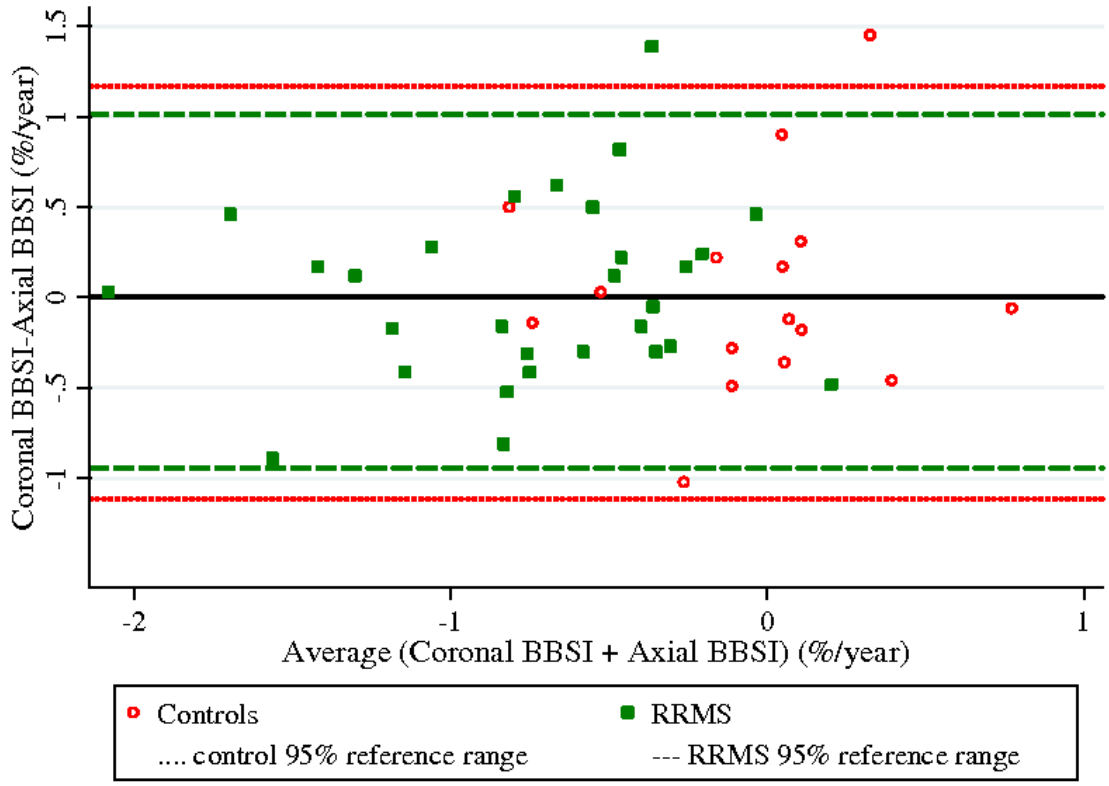


Figure 4-9 *Bland-Altman plot comparing BBSI brain atrophy rates quantified on coronal and axial T1-weighted volumetric images, in controls and subjects with RRMS. The reference ranges are the values within which 95% of the differences between atrophy measurements from coronal and axial images are expected to lie.*



There was no evidence of a difference in mean atrophy rates using the different methods in either controls (p=0.24) or subjects with RRMS (p=0.41). The generalised Pitman's test revealed a significant difference in the variance of atrophy rates in both controls and subjects with RRMS (both p<0.001). The SD of atrophy rates when BBSIs from coronal and axial images were averaged was the lowest of the five measures, and pairwise Pitman's tests revealed that it was significantly lower than the SD of measures obtained on the average images (using either coronal or axial ROIs) ($p \leq 0.001$). Consequently, sample sizes were estimated to be 22% and 18% higher (p<0.05) when quantifying atrophy on average images (coronal and axial ROI respectively) compared with averaging BBSI data from the two acquisitions (95% CI 8% to 38% higher when using the coronal ROI, and 4% to 35% higher when using the axial ROI). Based on pairwise Pitman's tests, no significant difference was observed in the variance of measures from either the single coronal or axial images and the average image, or between the coronal and axial images, although the latter had marginally less variance in rates.

Figure 4-10 *Intensity profiles of an average, axial and coronal image from one subject. A one-dimensional intensity profile through the brain on each image was generated by drawing a line one voxel high between the points A and B (top image), and normalising the intensity value of each voxel in this region to the mean whole brain intensity.*

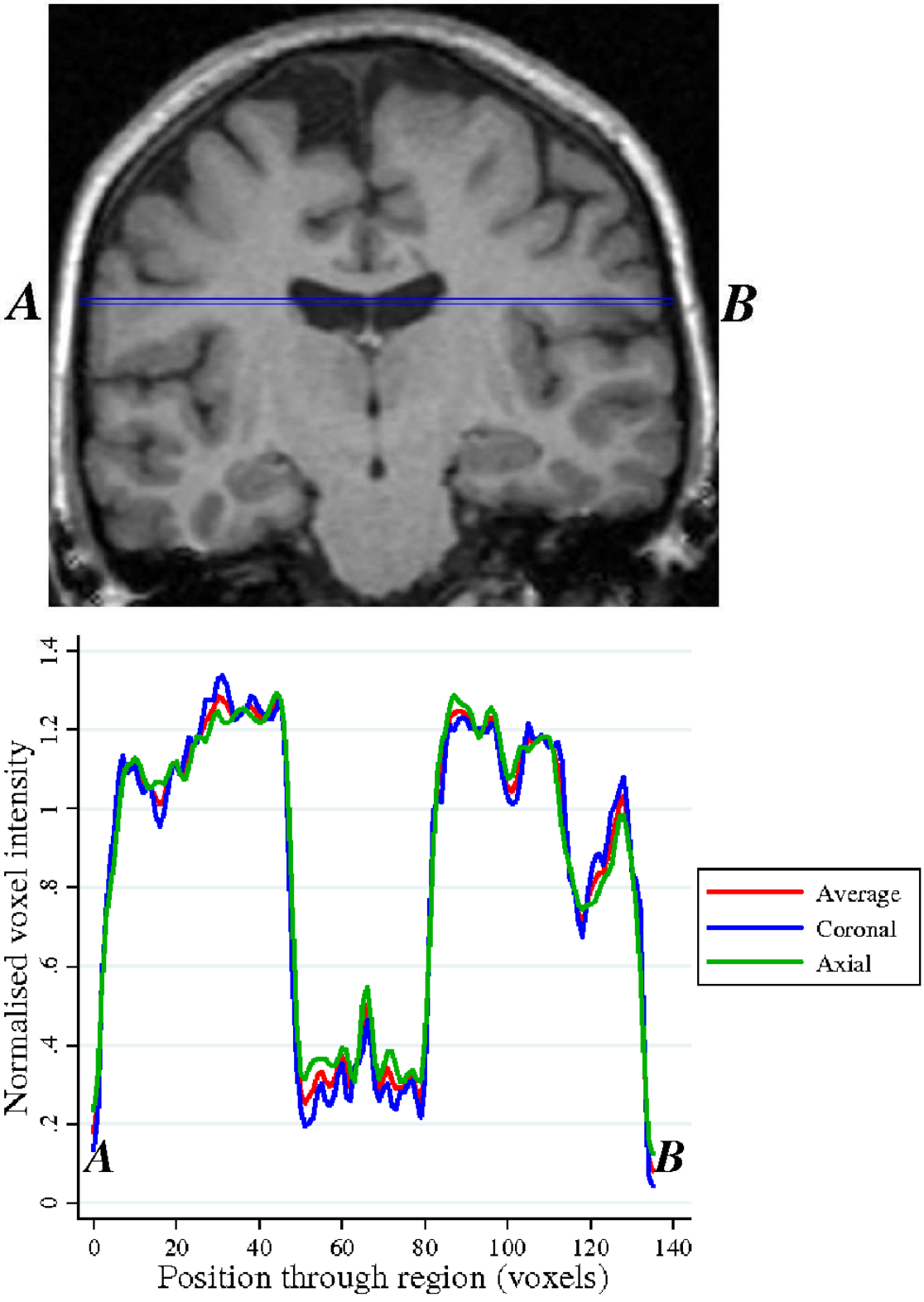


Figure 4-11 *BBSI brain atrophy rates quantified on average, axial and coronal T1-weighted volumetric images, in controls and subjects with RRMS.*

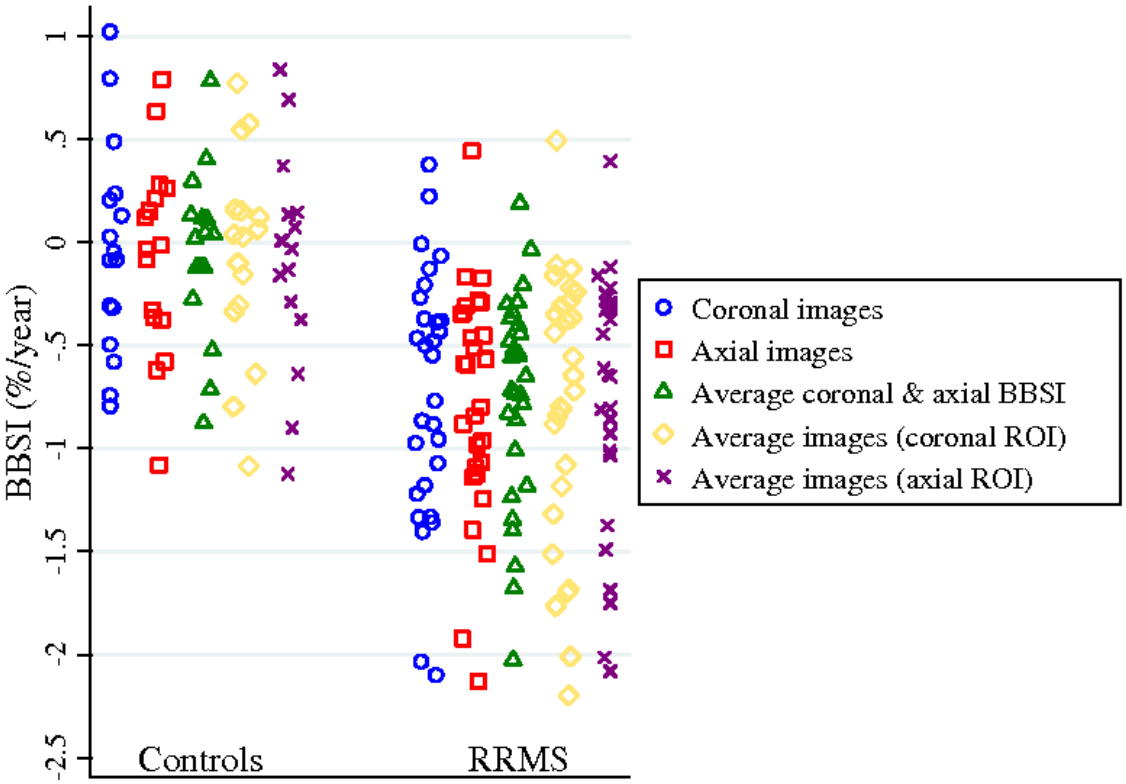


Table 4-4 *BBSI mean brain atrophy rates quantified on average, axial and coronal T1-weighted volumetric images, in controls and subjects with RRMS.*

	Mean (SD) atrophy rate (% year^{-1})	
	Control	***RRMS***
Coronal	-0.04 (0.52)	-0.73 (0.61)
Axial	-0.07 (0.48)	-0.76 (0.54)
Average BBSI	-0.05 (0.40)	-0.74 (0.52)
Average image (coronal ROI)	-0.05 (0.50)	-0.76 (0.65)
Average image (axial ROI)	-0.08 (0.50)	-0.80 (0.66)

4.4.4 Discussion

The aim of this study was to perform a direct comparison of the BBSI when applied to two different T1-weighted volumetric images and to investigate whether averaging data could improve the precision and power of atrophy measurement using the BBSI. No significant difference was observed between atrophy rates obtained on the two volumetric images, and it was shown that averaging repeated BBSI measures can decrease the variance of atrophy measurement by the BBSI. Interestingly, this simple strategy was superior to averaging MR images prior to atrophy measurement, and significantly increased statistical power.

Although the two images analysed in this study were acquired with similar protocols on the same scanner, small differences in their contrast and in the "sharpness" of edges was apparent, possibly due to the differences in the orientations of voxel dimensions and partial volume effects (Figure 4-8). However on average the SNR was greater in axial images compared with coronal images. Although within individual patients differences between atrophy rates obtained on coronal and axial images were apparent, there was no systematic bias between average measures from the two, suggesting that the differences are likely to have been due to noise from either the MR hardware or the measurement method. This is advantageous if the technique is to be used in future multicentre studies where differences in acquisition may occur between centres. As mentioned previously however, one of the limitations of this study is that the two images compared were acquired on the same scanner. To fully investigate the robustness of the BBSI, direct comparison of measures from images acquired on different makes and models of scanner should be performed. It is noteworthy that the axial acquisition had marginally less variance in measures. This is likely to have been due to the slightly larger voxel sizes of this acquisition which led to an increase in SNR. Small increases in voxels sizes for volumetric acquisitions may be a useful method by which measurement precision could be increased and should be investigated further.

Averaging the coronal and axial images significantly increased SNR relative to either individual acquisition. Despite the increased SNR there was no significant increase in the precision of atrophy measurement by the BBSI when applied to the average image compared with single images (coronal or axial), and the standard deviation was actually increased in RRMS patients. One reason for this finding may be due to the images having

been acquired in different planes. Correction for gradient distortions performed at the time of scanning is applied in-plane, but is not necessarily effective in removing through-plane distortion. Therefore distortion may have been different between the two images which may have led to more blurring at the edges of the brain and "cancelled out" any gain in signal that may have been attained through averaging. It may also have been due the limited number of scans available for averaging. A study that investigated the effectiveness of averaging up to 20 T1-weighted volumetric images (spoiled GRASS sequence, 1.0 x 1.0 x 1.0mm voxels) in creating enhanced MR images found evidence from intensity profiles that noise was reduced and that the GM/WM intensity crossings were much clearer on the average images (Holmes *et al*., 1998). It was also noted in this study that the benefit of averaging was apparent with as few as five images, but became noticeably greater as the number of images contributing to the average increased. It may be that average images have a greater impact on measures of regional or tissue-specific atrophy, where the contrast between GM and WM is vital.

The benefits of averaging repeated MR images and increasing measurement precision must be balanced against the time and expense of acquiring multiple MRI. As there was no significant difference in the variance of atrophy rates from coronal, axial and average images this implies that there is unlikely to be any substantial benefit from acquiring multiple images for averaging. Simpler, more cost- and time-efficient approaches may improve measurement precision. For example, in the study by Holmes *et al.* (Holmes *et al*., 1998), voxels were resampled to $0.5mm^3$ which may have improved measurements due to a reduction in partial volume effects. In addition, the current study found that the greatest measurement precision was achieved when two repeated BBSI measures were averaged; the variance was significantly lower for this measurement than for atrophy rates obtained on average images and this method consequently had greater statistical power. This result also suggests that noise introduced by the BBSI measurement technique may be greater than noise from the MR hardware.

One of the limitations of this study is that brain regions were not resegmented on the average image following its generation. Another study that investigated the use of principal component analysis applied to multispectral data to generate a composite image found that segmentation of images was superior on the composite image compared with single image segmentation (Zhang *et al*., 2001b). However the BBSI should be robust to

small differences in the segmentation and no significant differences in atrophy rates were observed between average images using the coronal or the axial ROIs.

In summary, this study has shown that the BBSI appears to be robust to small differences in acquisitions, and obtaining more than one scan at a study assessment for image averaging is not likely to be advantageous when quantifying the BBSI. Whilst averaging atrophy measurements by the BBSI increased measurement precision, generating average composite images did not improve measurement precision relative to single image analysis, despite significantly increasing SNR relative to single images. However this study has only investigated the averaging of two particular sequences; BBSI measurement on different 2D or 3D acquisitions may benefit from image averaging. Moreover, this process may be more advantageous when using other image analysis methods or when looking at regional brain areas. This study suggests that data averaging may be a simple approach to increasing the statistical power of brain atrophy measurement by the BBSI.

4.5 Chapter conclusions

This chapter describes automated brain atrophy measurement using the BBSI and implementation of procedures designed to optimise atrophy measurement using this technique. Whilst some optimisation strategies might be complex or labour intensive, it has been determined that using simple widely available techniques, the precision and sensitivity of measurements can be significantly increased. Correction of differential bias was shown to reduce the variability of atrophy rates and suggests that DBC should be performed routinely, even when there appears to be no disparity in intensity between images on visual inspection. Altering the window parameters for BBSI quantification was shown to increase measurement sensitivity. For large cohorts determining the optimal parameters with which to run the BBSI should be performed, as it can be fully automated and may increase statistical power. It has also been shown that obtaining multiple images at a given assessment to allow image averaging is not likely to improve statistical power significantly, and will increase the time required for both acquisition and post-processing. Whilst it may not be practical to analyse a dataset more than one time in large studies, for small exploratory studies averaging repeated measurements may be useful to increase the power of the study. Further investigation into the relative advantages of different acquisitions, specifically 3D images with larger voxel sizes (given the lower variance of measures on the axial image with larger voxels than the coronal), should be performed.

5 Analysis and optimisation of brain atrophy in multiple sclerosis using SIENA

5.1 Chapter introduction

As described previously, SIENA (Smith *et al*., 2002) is an automated algorithm, part of the FSL toolkit (www.fmrib.ox.ac.uk) (Smith *et al*., 2004), that allows the direct quantification of brain atrophy from serial MR images. It has been applied in numerous studies of MS, investigating brain atrophy in people who have experienced a CIS (Paolillo *et al*., 2004) and patients with RRMS (Richert *et al*., 2006), SPMS (Rovaris *et al*., 2003) and PPMS (Stevenson *et al*., 2002). Studies have shown brain atrophy rates ranging from -0.3 to -1.4% $year^{-1}$ (Filippi *et al*., 2004; Richert *et al*., 2006; Rovaris *et al*., 2003; Stevenson *et al*., 2002). As shown in Chapter 4 however, optimisation of automated methods for a subject group and scan acquisition protocol may increase the sensitivity and precision of measurement.

Few of the studies investigating brain atrophy in MS using SIENA have applied the technique to 3D volumetric MRI, although the technique is reported to be robust to different MR acquisitions. One of the most likely reasons for the lack of studies applying SIENA to 3D images is that if the initial automated brain extraction is not accurate, manual editing will be time-consuming for acquisitions with over 100 slices. The first aim of this experiment was therefore to determine whether accurate brain segmentation could be generated on 3D volumetric images by BET, which constitutes the initial stage of the SIENA algorithm. BET automatically removes skull and non-brain regions from the image, and certain algorithm parameters can be altered by the operator which will alter the resulting extracted region and can be used to alter the accuracy of segmentation. Like the BBSI, the brain ROI acts only as a guide to the brain region. Edge detection methods are used to determine the true brain boundary within this ROI, from which atrophy is calculated by comparing intensity gradient profiles between images (intensity inhomogeneity correction is performed as part of the algorithm). In theory this means that atrophy measures by SIENA are relatively protected from small errors in brain delineation, and should be less sensitive to these errors than methods that derive volumes from brain delineation. This chapter therefore also investigates whether a propagated template (average) brain region, which will not be entirely accurate for each subject, leads

to comparable atrophy rates by SIENA from volumetric MRI to those generated when individual brain masks are used.

Finally, as few studies have applied SIENA to volumetric acquisitions in MS, this chapter examines whether atrophy measurements by SIENA are comparable with those obtained on standard T1-weighted CSE imaging, the acquisition that has most commonly been utilised in previous studies.

5.2 Optimisation of brain extraction for a volumetric acquisition and investigation into the use of brain templates for atrophy measurements using SIENA

5.2.1 Introduction

As described previously, BET forms the initial stage of the automated SIENA algorithm (Smith *et al.*, 2002; Smith, 2002). Unlike the brain extraction used for the BBSI, BET does not attempt to model the brain surface at the finest level following gyri and sulci. Instead a tessellated surface mesh of triangles is deformed iteratively until a smooth surface fits the brain. Validation against manual methods have shown BET to be accurate for T1-weighted images although it was thought that the brain boundary was slightly overestimated by approximately one voxel (Smith, 2002). In addition, the validation study demonstrated that BET was robust to differences in image slice thickness (images with slices thicknesses ranging from 0.8-6mm were tested) and variations arising from differences in scanners (scans from 15 different scanners were tested, including 1.5T and 3T) (Smith, 2002).

With improvements to MRI scanner software and hardware decreasing scan times and improving image quality, and the development of new techniques for the analysis of high spatial resolution MRI, it is now common to acquire T1-weighted 3D volumetric images with slice thicknesses of 1.5mm or less. Despite validation of BET for images with thinner slices, few studies investigating brain atrophy in MS have applied SIENA to volumetric images. This may be because other factors influencing the appearance of an image may subsequently affect the performance of BET, such as individual scan parameters, resulting in scan contrast, intensity non-uniformity, and chemical shift and susceptibility-related artefacts. Although reducing the variability of atrophy measurements is important, high levels of accuracy are also desirable. Therefore prior to further processing by SIENA, manual editing of the images may be required, which can

be labour intensive for volumetric sequences. BET allows a number of parameters to be altered by the operator, and these can be used to optimise the brain extraction in an automated manner. Given the limited information available on the application of BET to 3D volumetric images in MS, the first aim of this study was to determine the accuracy of brain extraction using this tool on the set of 3D volumetric images being used in this thesis.

As slice-by-slice manual editing of sub-optimal brain regions can be labour intensive, it would be useful to develop methods which allowed accurate, robust and automated segmentation of the brain on any acquisition. Based on the hypothesis that SIENA should be relatively insensitive to small errors in brain segmentation, the second aim of this study was to address whether a standard template (average) brain region could be registered to images, and processed by SIENA, to give comparable brain atrophy rates to those obtained using accurate subject-specific brain segmentations.

5.2.2 Methods

Subjects and MR imaging

Sixteen control subjects (seven male, mean age 35.1 years (SD 6.3)), 42 subjects with CIS (14 male, 34.4 years (SD 6.3)) and 30 subjects with RRMS (eight male, 37.5 years (SD 7.4)) were included in the study. All subjects had coronal T1-weighted volumetric MRI (voxel dimensions 0.94 x 0.94 x 1.5) at baseline and approximately one year follow-up (mean follow-up time 1.16 years (SD 0.24)), according to the standard protocol described in Chapter 3.3. Images were reoriented to axial orientation, as SIENA is conventionally applied to axial images, whilst maintaining all voxel dimensions.

MRI analysis

BET was applied to all images using default parameters and the resulting brain masks inspected for accuracy. If the regions were not considered to be sufficiently accurate BET algorithm parameters were subsequently altered in order to optimise the automated brain extraction and obtain the most accurate brain region. The three parameters that could be altered were: i) the fractional intensity threshold, which leads to the overall segmented brain becoming larger or smaller, ii) the threshold gradient, which adds a gradient to the fractional intensity threshold leading to a larger overall segmented brain at the bottom of the image and smaller at the top, or vice versa, iii) the co-ordinates for the centre of the

initial brain surface sphere, which should lie approximately central in the brain. To estimate the co-ordinates for the centre of the brain, the MNI-152 standard space reference image (www.bic.mni.mcgill.ca/cgi/icbm_view) (Mazziotta *et al*., 2001) was registered to each image, using the FLIRT registration package (see Chapter 3.4.2) (12dof), maximising the normalised mutual information (Studholme *et al*., 1999). The brain segmentation available in FSL for this reference scan was transformed to each of the individual scans by applying the corresponding transformation parameters obtained from image registration, thereby creating an approximate brain mask on each scan in its native space. Using this approximate template brain mask, the co-ordinates of the centre of gravity of this mask were calculated and used as the brain centre. All optimised BET brain masks were visually inspected for accuracy and manually edited where required.

In order to generate template brain regions on each image, the same method used to apply an approximate brain mask to each image for estimation of the co-ordinates for the centre of the brain was applied, i.e. the MNI-152 standard space reference scan was registered to each image and the brain segmentation on this average scan was transformed to each individual scan by applying the corresponding transformation parameters. It should be noted that generation of the template mask on each image uses the same method as the "*betpremask*" option available in the FSL software package. It has been suggested that *betpremask* can be applied prior to BET, if BET alone is not producing optimal results. It was felt however, that the shape of the brain ROI used by betpremask was not always optimal for all images and often excluded inferior regions of the temporal and frontal lobes. Therefore the MNI-152 standard space brain template was manually edited to improve its accuracy before using an in-house script to perform the same process as *betpremask*. Each individual image was visually assessed to ensure that the template brain mask had been successfully transformed.

SIENA was run on all serial image pairs using i) the optimised BET brain mask, and ii) the template brain mask. PBVC was output for each subject using each mask and corrected for scan interval. The accuracy of registration of baseline and repeat images was checked as was the colour overlay showing regions of atrophy.

Statistical analysis

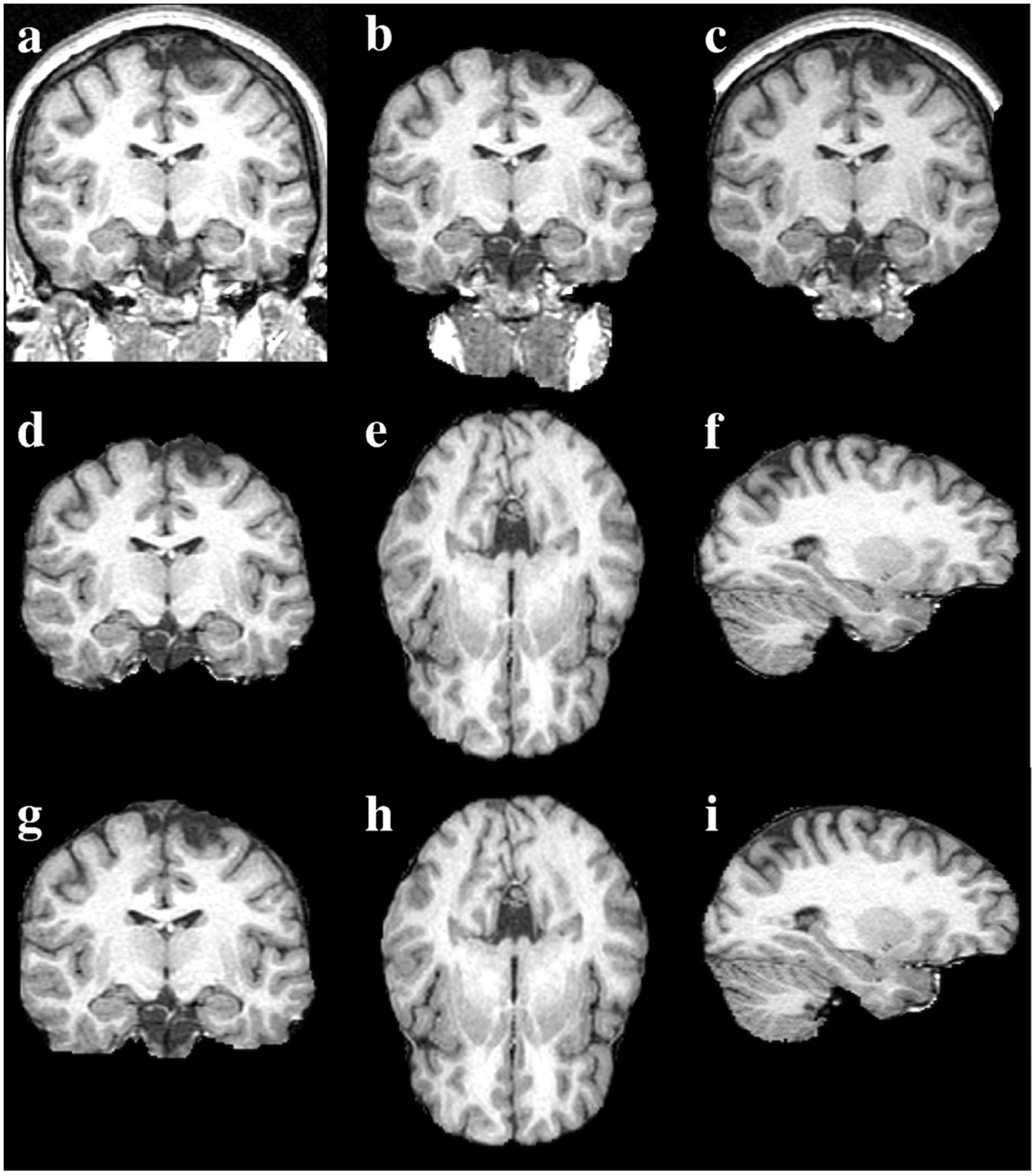
Mean (SD) atrophy rates (% year^{-1}) were determined for control, CIS and RRMS subject groups when using subject-specific brain masks and the template brain mask. Within each subject group a paired *t*-test and Pitman's test was performed to determine if there were significant differences in the mean and variance of atrophy rates respectively, between measures obtained using subject-specific brain masks and the template brain mask. In addition, a Bland-Altman plot was created to assess the differences between atrophy rates obtained when using the two different brain masks. As in Chapter 4, relative statistical power for atrophy measurements obtained using the different masks was calculated if there was found to be significant differences in either the mean or variance of measures.

5.2.3 Results

Figure 5-1 shows examples of the brain regions obtained using BET in one subject from the cohort. Default parameters gave inaccurate brain regions which were not improved by altering the intensity threshold and threshold gradient (Figure 5-1a to c). By specifying the co-ordinates of the centre of the brain for the initial brain surface sphere, accurate brain extraction was achieved for all images (Figure 5-1d to f). There were no failures when creating the template brain region but small errors were apparent on all scans (Figure 5-1g to i), mainly exclusion of the edges of the inferior and lateral temporal lobes, anterior frontal lobes and inferior cerebellum. In addition there was some inclusion of dura in the superior regions of the head in some subjects.

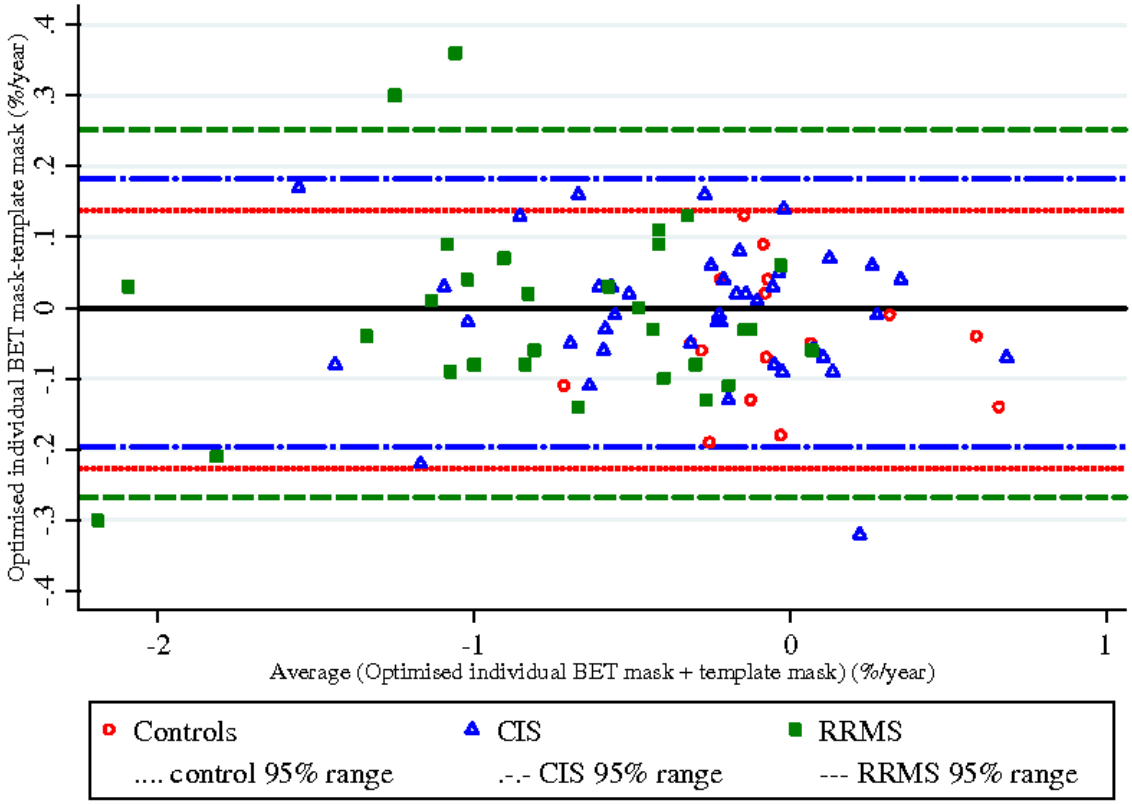
In controls mean atrophy rates were -0.07% year^{-1} (SD 0.34) using the optimised individual BET brain mask, and -0.03 % year^{-1} (SD 0.34) using the template mask. Similarly, in CIS subjects mean atrophy rates were -0.31% year^{-1} (SD 0.48) with individual masks and -0.30% year^{-1} (SD 0.49) with the template mask, whilst in RRMS subjects they were -0.78% year^{-1} (SD 0.59) and -0.77% year^{-1} (0.57) respectively. Looking at the Bland-Altman plot (Figure 5-2) there was no apparent bias in measurements obtained using the two different masks, and the 95% reference range (the values within which 95% of values are expected to lie) was relatively narrow with all differences below 0.4% year^{-1}. This was confirmed by quantifying the mean difference (individual BET mask – template mask) between atrophy rates, and paired *t*-tests gave no evidence that there was a significant difference.

Figure 5-1 *Examples of the BET brain extraction using different parameters, and the template brain extraction from a coronal T1-weighted volumetric image. a) original image, b) BET default parameters, c) changing the fractional intensity and gradient thresholds, d)-f) specifying the co-ordinates for the centre of the initial brain surface sphere of BET, g)-i) using a template brain mask.*



Mean differences were -0.04% year^{-1} in controls (SD 0.09, 95% CI -0.09 to 0.01, p=0.08), -0.01% year^{-1} in CIS subjects (SD 0.10, 95% CI -0.04 to 0.02, p=0.62), and -0.01% year^{-1} in subjects with RRMS (SD 0.13, 95% CI -0.06 to 0.04, p=0.75). In addition, no significant differences were seen in the variance of measures calculated using the two different masks (controls p=0.94, CIS p=0.42, RRMS p=0.56).

Figure 5-2 *Bland-Altman plot comparing SIENA brain atrophy rates quantified using optimised individual BET brain masks and a template brain mask on T1-weighted volumetric images, in controls and subjects with CIS and RRMS. The reference ranges are the values within which 95% of the differences between atrophy measurements using the two different masks are expected to lie.*



5.2.4 Discussion

This study investigated the performance of BET on T1-weighted volumetric images, and the relative brain atrophy rates obtained using SIENA when applying subject-specific accurate brain masks and a template brain mask. It was shown that brain extraction can be achieved on these images using BET, without the need for manual editing. In addition, it was demonstrated that using an average template brain mask, similar brain atrophy rates

are quantified by SIENA suggesting that the technique is robust to small errors in the brain extraction.

Although using default parameters BET gave poor results on the images included in this study, it was found that by specifying the co-ordinates for the centre of gravity of the head from which the deformable model is initiated, accurate results could be achieved on all images. The most likely reason for this is that images analysed in this study were acquired in the coronal plane, and the brain was not positioned centrally within the FOV (when viewing it in the coronal plane, i.e. neck was included in the image and the brain therefore was positioned higher). If the deformable model is initiated at the centre of the FOV, this consequently results in large regions of non-brain, inferior to the brain, being included in the extraction. By specifying the approximate co-ordinates for the centre of the head the deformable model can achieve accurate brain segmentation. Indeed a study published in 2007 which included volumetric images that contained a lot of lower non-brain matter devised a simple script that re-ran BET several times, each time intialising the brain centre estimation using the centre of gravity found on the previous iteration (Smith *et al*., 2007). In this way accurate brain extraction was achieved for downstream processing by SIENA. Another study that used BET to process T1-weighted volumetric MPRAGE and SPGR images (both $1mm^3$ voxels) similarly found that accurate brain extraction could be achieved by altering BET parameters (Clark *et al*., 2006). In this case the threshold for extraction was optimised for the different pulse sequences. It seems that the optimisation of BET will depend to a large extent on the acquisition in question.

Use of a registration template is advantageous, as it is fully automated, reproducible, can be applied to any acquisition, and the cost function (normalised mutual information) should not be influenced by image inhomogeneities, movement or lack of boundary definition (Maes *et al*., 1997), making it highly reproducible providing that the registration technique is robust. It also avoids labour intensive manual editing or scan specific optimisation (as was used with BET on 3D acquisitions). Although templates may represent a trade-off between accuracy versus speed and reproducibility, when using a standard template brain mask in this study, atrophy rates closely approximated those obtained when the individual and accurate BET brain masks were applied. As already mentioned, these results appear to suggest that SIENA may be robust to small errors in brain segmentation. A study that investigated intercentre agreement of brain atrophy

measurement in MS using SIENA on 2D images found that manual editing of brain regions obtained by BET did not significantly alter mean atrophy rates (Jasperse *et al*., 2007b), suggesting that downstream processing is not affected when some non-brain regions are included in the initial brain mask. This is in agreement with the findings of the experiments described in this chapter. However the variance of measurements within a centre was greater compared with when manual editing of regions was performed which is in contrast to the present study.

Although a standard brain mask was used as a template, this can lead to inaccuracies due to the normal variation in human brain between subjects. One approach that has been investigated that could lead to more accurate segmentation is the creation of a specific template for a given dataset. For example a template specific to patients with RRMS or controls could be created and applied, or a family of brain templates could be collated which are then searched to find the "best" template for a given image, as proposed in one study (Wu *et al*., 2007a). Non-linear registration of a template brain mask to images may also improve the accuracy of resulting individual brain masks.

From this study, it appears that high resolution 3D acquisitions should be considered for future brain atrophy studies in MS as (i) automated brain segmentation can be achieved using BET, and (ii) SIENA appears to be robust to small segmentation errors in the brain mask, which may not be true on images that have lower resolution. The use of template brain masks on a larger cohort of subjects with T1-weighted volumetric images from different MR scanners should be performed.

5.3 A comparison of SIENA performance on "3D" volumetric acquisitions and "2D" spin echo acquisitions

5.3.1 Introduction

Typically SIENA has been applied to T1-weighted "2D" CSE images with 3mm or 5mm thick contiguous slices (Rovaris *et al*., 2003; Stevenson *et al*., 2002), in part because the acquisition is acquired for other analyses, but also because the number of slices comprising the image is limited, minimising manual editing when required. It was shown in the last experiment, that an automated and accurate brain extraction can be achieved on volumetric T1-weighted acquisitions using BET. If analysis time is not increased, this finding may promote the acquisition of 3D volumetric images for future studies using

SIENA. One might expect more reliable atrophy measurements to be obtained on images with higher resolution (i.e. 3D volumetric images as opposed to 2D CSE images), where partial volume effects will be minimised and sensitivity to detect small tissue volume changes may be improved. However little investigation comparing atrophy quantification by SIENA on the two sequences has been performed.

In the original validation of SIENA by Smith *et al.* (Smith *et al*., 2002), it was found that image slice thickness had no effect on longitudinal atrophy measures using SIENA. Yet this observation was based on analysis of control subjects who had been scanned over a short interval and in whom no atrophy was expected. It is unclear whether in patients in whom brain volume loss is occurring, longitudinal atrophy measurement by SIENA is influenced by image slice thickness and whether using 3D T1-weighted images provides better results than 2D images in relation to measurement precision, sensitivity, processing speed and reliability. This aim of this experiment was to evaluate the performance of SIENA on 2D and 3D T1-weighted acquisitions in patients with RRMS in whom brain atrophy was occurring, and control subjects.

5.3.2 Methods

Subjects and MR imaging

Twenty-nine subjects with RRMS (seven male) and 14 normal healthy control subjects (five male) with MRI at baseline and one year follow-up (mean 1.1 years (SD 0.1)) were identified for the study. At baseline the MS group had a mean age of 37.3 years (SD 7.4) whilst controls had a mean age of 35.3 years (SD 7.4). Mean disease duration in patients was 3.0 years (SD 0.8) and disability was mild, with a median EDSS of 1.5 (range 0-3).

T1-weighted 2D CSE and coronal 3D FSPGR images were obtained using the acquisition parameters described in Chapter 3.3.2. Voxel dimensions were 0.94 x 0.94 x 5mm^3 for the CSE image and 0.94 x 0.94 x 1.5mm^3 for the FSPGR image. Both acquisitions were acquired on the same day.

MRI analysis

Three-dimensional volumetric images were reformatted to axial orientation, whilst maintaining all voxel dimensions. Individual brain masks were generated for each 2D and 3D image using the optimised version of BET described in Chapter 5.2, which required

specification of the co-ordinates of the approximate centre of the brain for the initial brain surface sphere. Results were visually assessed and in the case of 2D images a judgement was made regarding the need for further optimisation of BET. Alteration of the intensity threshold, threshold gradient, or manual editing was performed as necessary on the 2D images in order to obtain an accurate brain mask on each image. SIENA was subsequently run on each 2D and 3D serial image pair, which registered the repeat brain to the baseline brain prior to atrophy quantification. All resulting PBVCs were annualised.

Statistical analysis

When assessing the results several factors were considered. Firstly the level of automation and operator time required to obtain an accurate brain mask on a given image was considered. It was also noted if there were any failures in either creating the brain masks or with SIENA. Mean (SD) atrophy rates in controls and patients with RRMS were calculated on the two acquisitions. Atrophy rates obtained on the 2D and 3D acquisitions were compared pairwise by subject using a Bland-Altman plot and paired *t*-tests to determine if there was bias between the two acquisitions. Pitman's test was used to determine if there was a significant difference in the variance of measures obtained on the two acquisitions. The mean (SD) difference between 2D and 3D atrophy rates was also determined.

Investigation into the relative statistical power of 2D and 3D acquisitions for monitoring brain atrophy was performed. As in Chapter 4.2 the CV (σ/μ) of the two methods was calculated, and the square of their ratio was determined to indicate the relative number of patients required to detect a treatment effect using the two different acquisitions.

5.3.3 Results

Brain masks obtained on the 2D images using the same optimised version of BET as for 3D images were sub-optimal, including non-brain regions mainly in the area of the eyes (Figure 5-3a). Due to the nature of the errors it was decided that altering the intensity thresholds would not significantly improve brain extraction, so manual editing was performed on all images (taking approximately 20 minutes per subject). Creation of the brain masks on the 3D images using the optimised version of BET was fully automated and none of the subjects failed. An example of the resulting brain mask is shown in Figure 5-3b.

Figure 5-3 *Example of the brain extraction (outline) obtained using an optimised version of BET on a) a T1-weighted 2D spin echo image and b) a T1-weighted 3D volumetric image.*

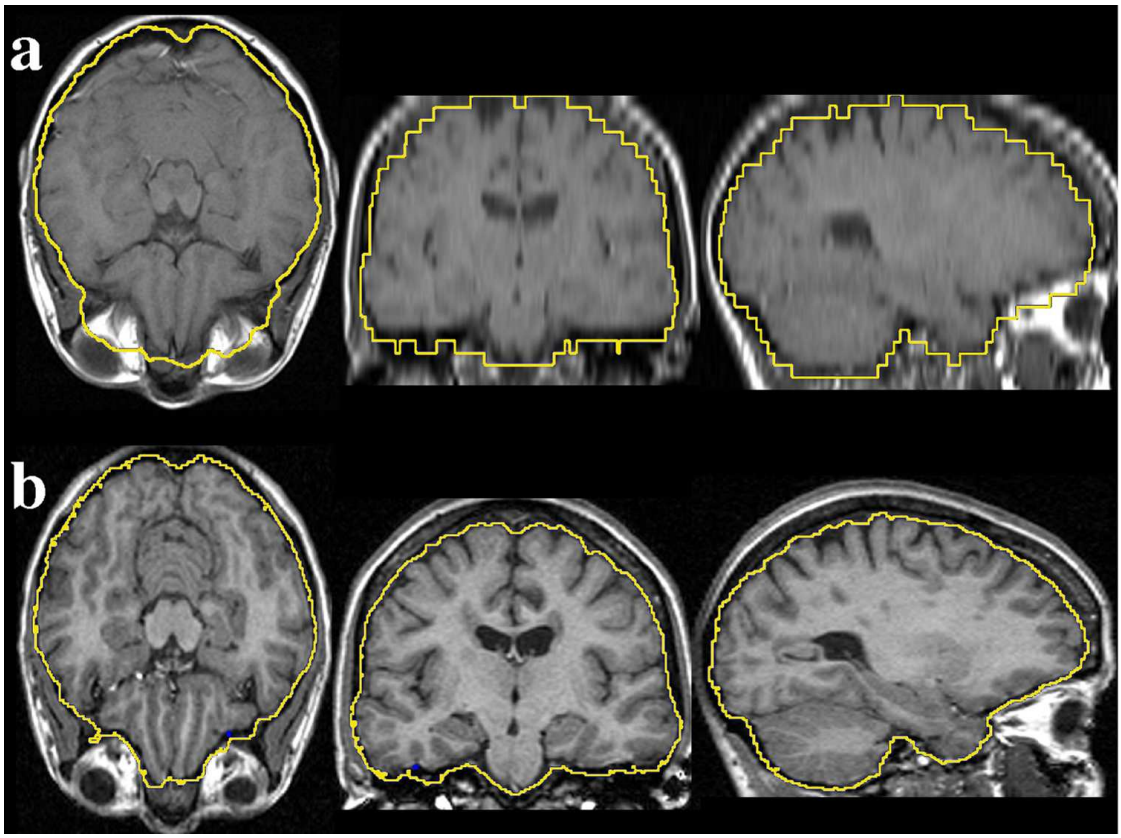


Figure 5-4 gives an overview of the brain atrophy rates obtained in controls and patients with RRMS on each of the acquisitions whilst Figure 5-5 shows the Bland-Altman plot comparing brain atrophy rates obtained on the two acquisitions. For most subjects agreement between the two measures was reasonable, however three patients, who have greater atrophy, show much larger atrophy rates on the 3D compared with the 2D acquisition, suggesting there may be some bias in measures. As can be seen on the plot, differences up to around 1% year^{-1} between 2D and 3D acquisitions were observed, and the range of differences within which 95% of subjects are estimated to lie is wide (shown on the Bland-Altman plot).

Figure 5-4 *SIENA brain atrophy rates quantified on 2D spin echo and 3D volumetric acquisitions, in controls and subjects with RRMS.*

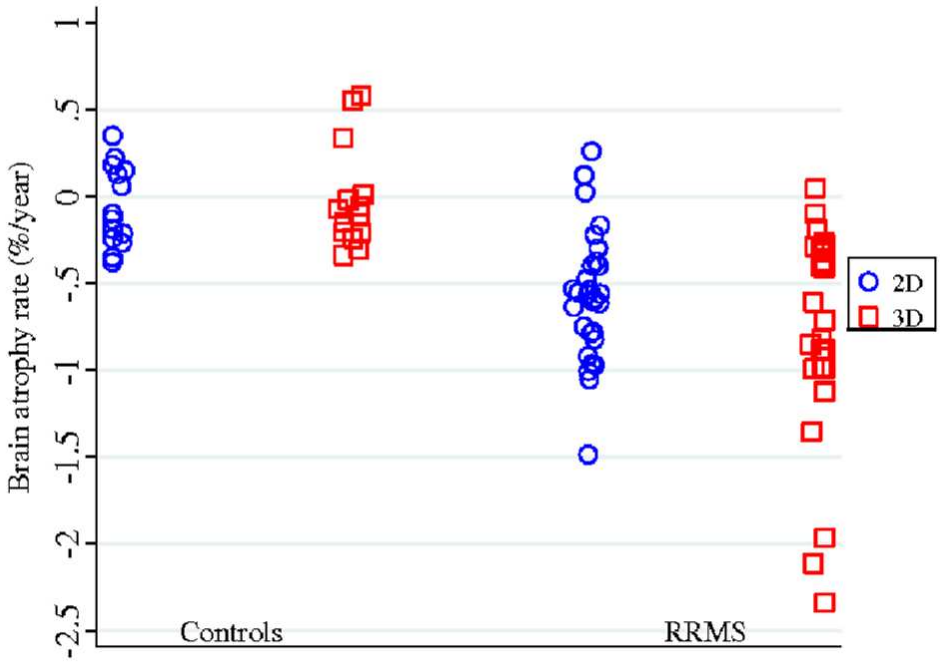
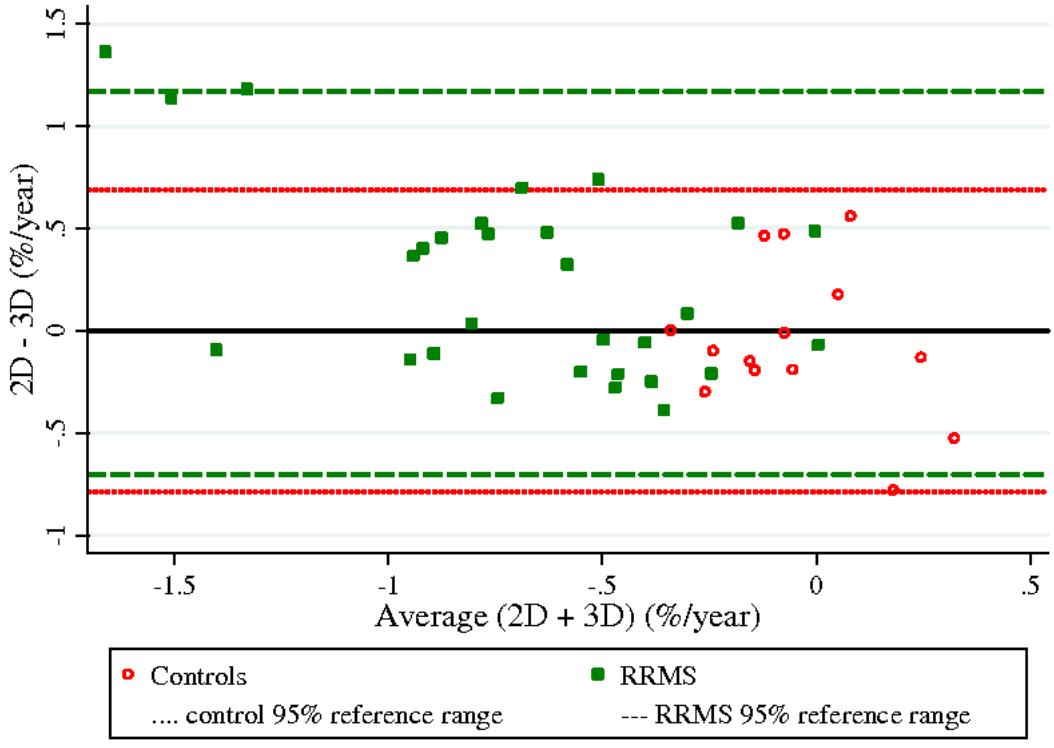


Figure 5-5 *Bland-Altman plot comparing SIENA brain atrophy rates quantified on 2D spin echo and 3D volumetric images, in controls and subjects with RRMS. The reference ranges are the values within which 95% of the differences between atrophy measurements from 2D and 3D images are expected to lie.*



In controls, there was no evidence that the atrophy rates differed between 2D and 3D acquisitions with mean (SD) atrophy rates of -0.07% year^{-1} (0.23) and -0.02% year^{-1} (0.30) acquired on 2D and 3D images respectively. The mean difference (2D-3D) in atrophy rate was only -0.05% year^{-1} (SD 0.38, 95% CI -0.26 to 0.17, p=0.63). Although the variance of atrophy measures was marginally smaller on the 2D acquisition compared with the 3D acquisition, this difference was not significant, p=0.37.

In RRMS patients the atrophy rate quantified on 2D and 3D images agreed less than in control subjects. Mean atrophy rates were -0.57% year^{-1} (SD 0.36) and -0.80% year^{-1} (SD 0.58) acquired on 2D images and 3D images respectively. The mean difference (2D-3D) in atrophy rate was statistically significant at 0.24% year^{-1} (SD 0.48, 95% CI 0.06 to 0.42, p=0.012). The variance of measures was significantly smaller using 2D acquisitions in patients with RRMS (p=0.003). However the increased mean and variance with 3D acquisitions were of a similar magnitude and hence sample size requirements using the two acquisitions were not significantly different. Sample sizes to detect a particular proportionate difference were estimated to be 13% smaller (not significantly different) using 2D acquisitions (95% CI 45% reduction to 28% increase).

5.3.4 Discussion

This study has been able to analyse directly the differences in brain atrophy rate quantified on 2D and 3D MR acquisitions by SIENA, using same-day MRI scans of controls and subjects with RRMS. This is the first study to directly compare SIENA on different MR acquisitions in patients with a disease-related increase in brain atrophy. Some limitations of the study should be noted however. Firstly, this study was based on images that were acquired on one MR scanner, and therefore one should be cautious in generalising these results to scans acquired on other scanners or in multicentre studies. Secondly, the FOV for 2D images did not include full head coverage superiorly, with some skull and the tips of some gyri excluded (Figure 5-3). This may have biased results slightly in terms of the stability of both brain extraction and registration of 2D images. Edge detection by BET is to some extent based on an intensity gradient, and the lack of gradient when the tips of gyri are excluded may have influenced the segmentation. However, as can be seen from Figure 5-3, problems with the segmentation were primarily around the eyes. Registration of images by SIENA is optimised using the skull and therefore may also have been affected on 2D images. However as described by Smith *et*

al. (Smith *et al*., 2001b), a small amount of "missing skull" is not generally a problem, and in addition all registrations were acceptable when checked.

In the cohort described in this paper it has been shown that within individuals there may be differences observed between atrophy rates quantified on 2D and 3D acquisitions. In control subjects, in whom little atrophy is occurring, atrophy rates appear to agree on average. However a significant difference in atrophy rates acquired on 2D and 3D imaging was found in RRMS patients, a finding driven by the patients showing greater atrophy over the course of a year. Differences in measured atrophy rates such as these, which are in the order of up to 1% year^{-1}, will impact on the ability to compare results from studies using different acquisitions, as mean brain atrophy rates in RRMS have been shown to be in the order of only -0.9 to -0.5% year^{-1} (Fox *et al*., 2000b; Kalkers *et al*., 2002; Richert *et al*., 2006; Rovaris *et al*., 2003). However, as it is difficult to know the true atrophy rate for a subject, it is not feasible to say whether 3D acquisitions provide more accurate measures than 2D acquisitions.

In these patients it appeared that 3D acquisitions were more sensitive in detecting atrophy than 2D acquisitions. The mean atrophy rates found on 2D acquisitions in this study were marginally lower than those observed in previous studies however. A median atrophy rate of -0.7% year^{-1} (SD 0.9) was found in one study in RRMS where the acquisition consisted of 5mm thick slices (Rovaris *et al*., 2003). In two other studies of RRMS that applied SIENA to CSE acquisitions with 3mm thick slices, mean atrophy rates of -0.87% year^{-1} (SD 0.34) (Richert *et al*., 2006) and 0.9% over a nine month interval (SD 1.2) (Sormani *et al*., 2004) were found. This discrepancy may be due to differences in the cohorts, and disease duration and severity are likely to influence rates considerably. It may also have been due to the brain region obtained on the 2D acquisitions excluding the most superior tips of the gyri in some patients, which may have decreased the amount of atrophy detectable on 2D acquisitions. It should be noted that the variance of brain atrophy rates was significantly lower for 2D acquisitions than 3D acquisitions in RRMS subjects and hence the two acquisitions demonstrated similar statistical power. In other words the increased sensitivity and greater mean atrophy rate on 3D scans was offset by a greater interpatient variability in the rates of atrophy. The result is a neutral effect in sample size requirements for demonstrating a therapeutic effect on the rate of atrophy. However if the difference between 2D and 3D acquisitions is greater in patients with more atrophy,

identification of patients with higher atrophy rates for inclusion into treatment trials could provide smaller sample sizes if using 3D imaging, and this should be investigated further. One question that should be addressed in relation to the difference in variance between the two acquisitions is whether there is a systematic increase in measurement variability because 3D acquisitions take longer to acquire and may be more prone to movement artefact.

Other studies that have compared atrophy measurements on 2D and 3D acquisitions include an investigation into GMFs and WMFs which showed significant differences in the measurements from four different acquisitions (including 2D and 3D images) (Zivadinov *et al*., 2004b). The study concluded that the optimal pulse sequence for measurement was a 3D spoiled gradient echo acquisition. In addition, a cross-sectional study investigating measures of BPF found that 2D and 3D acquisitions were equally sensitive in distinguishing controls and MS patients (Sharma *et al*., 2004).

One advantage of the 3D acquisitions that was observed in this cohort was the ability to obtain brain regions using fully automated methods, whilst manual editing was required on all 2D acquisitions. This may be due to the lower image resolution with increased partial volume effects influencing segmentation on the 2D acquisition. A study comparing an automated brain extraction algorithm on 2D and 3D acquisitions from 52 MS patients also found that unreliable segmentations were obtained on 2D images, whilst on 3D images the segmentations were acceptable (Sharma *et al*., 2004). However it should be noted that in addition to spatial resolution, other factors may affect the appearance of MR images, and subsequently the performance of BET, such as the scan parameters, field strength, resulting scan contrast, intensity non-uniformity, and chemical shift and susceptibility-related artefacts.

Although not investigated in this study, one might expect the reproducibility of atrophy measures to be greater on 3D than on 2D acquisitions. However one study that investigated scan-rescan reproducibility showed worse reproducibility on 3D acquisitions compared with 2D acquisitions, when BPF was measured cross-sectionally using an automated algorithm (Horsfield *et al*., 2003). It would be useful to determine the scan-rescan reproducibility of longitudinal SIENA measures on 2D and 3D acquisitions to add to the current findings.

In conclusion, differences between atrophy measures using SIENA on 2D and 3D acquisitions exist and may be larger when atrophy is occurring. Comparisons between studies using SIENA on different acquisitions should be made with caution. However the variability of measures relative to the mean was similar for both acquisitions, suggesting that they have similar statistical power. It has been shown that it is possible to obtain brain regions using BET in a fully automated manner on 3D volumetric acquisitions, negating the need for labour-intensive manual editing. Given the similar statistical power of the two acquisitions, this suggests that further investigation into the application of SIENA to other 3D volumetric images and in other MS clinical subgroups who may have higher atrophy rates (e.g. SPMS and PPMS, rapidly deteriorating RRMS) should be performed.

5.4 Chapter conclusions

This chapter has investigated the performance of SIENA on T1-weighted volumetric MRI, as few studies have previously applied the technique to 3D images. Volumetric MRI is now commonly acquired in patients with MS and improvement to MR scanner hardware and software means that they can be easily and quickly obtained. This chapter has shown that automated brain extraction can be achieved on these images using an optimised version of BET, which will increase the reproducibility of measurements. In addition, atrophy measurement by SIENA was robust to small errors in brain extraction when a template brain mask was used. The use of such a method may be an important consideration for large studies where brain atrophy analysis is to be performed at multiple centres.

Differences appear to exist in atrophy quantification by SIENA on these 3D volumetric images and 2D CSE images to which SIENA has been applied more often. From this small study it seems that atrophy measurements should not be compared or combined between studies using different acquisitions. Although there may be advantages to using 3D volumetric acquisitions in future studies, given that automated brain extraction may be more robust, 2D acquisitions gave marginally greater statistical power, although this was not statistically significant. This may be more important in future studies designed to detect differences in atrophy rates between groups. One should be cautious in generalising these results to other 2D and 3D acquisitions however, as this study compared only one particular 2D acquisition and one particular 3D acquisition. A large number of different 3D acquisitions exist (and from different vendors) and it may be that 3D acquisitions with

larger voxels that could be acquired in a shorter acquisition time (and therefore be less noisy) would improve 3D performance. Further investigation into the relative atrophy measures from other 2D and 3D acquisitions should be performed in larger cohorts and other clinical subgroups to extend these findings.

SIENA has been shown to provide automated reliable brain atrophy measurements in patients with MS and controls. Investigation into the relative performance of this technique with other automated and semi-automated brain atrophy measurement methods should be performed.

6 A comparison of brain atrophy measurement techniques

6.1 Chapter introduction

The development of new disease-modifying treatments that may target the underlying causes of disability in MS, namely neuroaxonal damage, has increased the need to monitor disease progression as precisely (and accurately) as possible. Measurement of brain atrophy is thought to be a promising marker of this cerebral damage. A major motivation for research in this area is that identification of robust, sensitive and precise techniques to measure global or regional atrophy may aid in detecting effective treatments that slow disease progression.

In Chapters 4 and 5 it was shown that the BBSI and SIENA can be optimised for T1-weighted volumetric MRI and used to detect brain atrophy in patients with MS early in the course of their disease. Other methods commonly used to quantify atrophy rates are based on segmentation of ROIs on serially acquired MRI, and calculating the difference in ROI volumes over the scan interval. These ROIs include WBV, central cerebral brain volume, ventricular volume, GM and WM volumes (Dalton *et al*., 2006; Rovaris *et al*., 2000; Tiberio *et al*., 2005). A variety of software and algorithms used for ROI segmentation have already been described in Chapter 1.6. Whilst these methods may have been validated for atrophy measurement, they have been applied to diverse study cohorts making it difficult to determine their relative performance and ability to measure brain atrophy in patients with MS. The few studies that have compared atrophy measurement techniques have been limited, for example comparing only two techniques (Sormani *et al*., 2004), or comparing the same measurement (e.g. BPF) using different algorithms (Sharma *et al*., 2004; Zivadinov *et al*., 2005).

This chapter directly compares several commonly used methodologies for measurement of brain atrophy in MS, to investigate which might provide the most sensitive, precise and reliable measure of atrophy, and therefore which could be the most effective in monitoring disease progression in MS. In the first part of this chapter a direct comparison of the two registration-based methods, the BBSI and SIENA, is carried out. The second part of this chapter explores other methods that have been applied in brain atrophy studies, and are thought to be good markers of brain tissue loss.

6.2 A comparison of registration-based methods of brain atrophy

6.2.1 Introduction

The previous two chapters have confirmed that the BBSI and SIENA are sensitive to disease-related atrophy, and it has been shown that they can be optimised for a given cohort and particular MR acquisition to improve the accuracy, sensitivity and precision of measurement. As has already been described, both the BBSI and SIENA consist of algorithms that register serial MR images from the same individual and estimate brain volume change through detection of changes at the edge of the brain. However some differences between the two techniques exist, and a comparison of them is presented in Table 6-1. It should be noted that although a degree of operator interaction is required to obtain an initial brain segmentation for registration, BBSI brain atrophy quantification itself is calculated automatically.

Table 6-1 *A comparison of BBSI and SIENA methodology.*

Similarities	***Differences***	
	BBSI	***SIENA***
Automated registration-based technique	Scaling changes optimised using brain	Scaling changes optimised using skull
Brain segmentation required prior to registration and atrophy quantification	Semi-automated brain segmentation using MIDAS image analysis software	Automated brain segmentation using brain extraction tool
Atrophy quantification based on movement of brain edge	Quantification based on intensity differences between brain edges	Quantification based on distance moved by brain edge
Small segmentation errors should not affect quantification	May be affected by intensity changes from baseline to repeat image	Relatively insensitive to intensity changes

Although both these registration-based methods have been applied to MS subjects in the previous chapters, it is unclear how they compare to one another, and to differences in segmented brain volumes, in their ability to detect atrophy in subjects with early MS and CIS. A study published in 2006 did investigate the effect of registration on measurement of short-term (three months) brain volume change using different segmentation

techniques and SIENA. However no change in brain volume was detected over this period and no significant differences in volume changes were observed on registered and non-registered images when comparing results between techniques (Fritz *et al*., 2006).

The first objective of this study was to compare directly: i) segmented BVD on non-registered images ii) the BBSI and iii) SIENA, through quantification of brain atrophy rates over one year in subjects with CIS, early RRMS and controls. If brain atrophy is to be utilised in clinical trials, one might assume that an effective treatment would reduce atrophy rate to the level observed in control subjects, and therefore the ability of a technique to differentiate between control and patient groups may indicate the statistical power of a method. Moreover, it is important to determine how well any potential atrophy measure relates to disability. Detection of subtle degrees of atrophy early in the course of disease, particularly at the CIS stage, may aid diagnosis and prognosis of individual patients. The second objective of the study was therefore to compare the ability of each method to differentiate between patient groups and controls, and between patients with CIS who had developed MS at a three year follow-up compared with those who had not.

6.2.2 Methods

Subjects and MR imaging

This study included 83 subjects: 37 patients (14 males) presenting with CIS, 30 patients (eight males) with early RRMS (less than four years duration) and 16 control subjects (seven males) as described in Chapter 3.1. Thirty-four patients with CIS presented with optic neuritis, two with spinal cord syndromes and one with a brain stem syndrome, and patients ranged in age from 21 to 48 years at baseline (mean 35.0, SD 6.2). Patients with RRMS ranged in age from 26 to 56 years (mean 37.5, SD 7.4), and mean disease duration (from the first clinical episode) was 2.0 years (SD 0.8, range 0.5 to 3.8). Two patients were on beta interferon from baseline, whilst a further two patients started beta interferon treatment during follow-up. Controls ranged in age from 27 to 53 years (mean 35.1, SD 6.3). The EDSS (Kurtzke, 1983) was used to assess disability in patients at the time of baseline MRI and median (range) EDSS score was 1.0 (1.0 to 2.5) in patients with CIS and 1.5 (0 to 3.0) in patients with RRMS. A coronal 3D FSPGR sequence was acquired on all subjects at baseline and approximately one year later (mean 1.1, SD 0.2, range 0.9 to 1.8), according to the protocol described in Chapter 3.3. Baseline MRI was performed within 12 weeks of symptom onset in patients with CIS (mean 6.0, SD 3.4).

MRI analysis

(i) Segmented brain volume difference

Semi-automated segmentation of baseline and repeat brains was performed using MIDAS as described in Chapter 4.2.2, with correction of images for intensity inhomogeneity using N3 (Sled *et al*., 1998). Brain volume was determined from the segmentation within MIDAS and the baseline volume subtracted from the follow-up volume. Cerebral volume loss was expressed as a percentage of total baseline brain volume and annualised to give a global atrophy rate.

(ii) BBSI

The BBSI was applied to images as described in Chapter 4.2.2. DBC (Lewis & Fox, 2004) was applied at registration and the BBSI window parameters were set at the optimal values determined in Chapter 4.3 (I_1=0.65, I_2=0.45). Cerebral volume loss was expressed as a percentage of total baseline brain volume and annualised to give a global atrophy rate.

(iii) SIENA

SIENA was applied as described in Chapter 5.3.2. This required axial reorientation of all images, and the initial brain segmentation was acquired using the optimised version of BET, described in Chapter 5.2. PBVC calculated by SIENA was annualised to give a global atrophy rate.

Three year clinical assessment of patients with CIS

All but two of the patients with CIS (n=35, 12 males) had clinical assessment at three-year follow-up, when they were evaluated for a diagnosis of MS (progression from CIS) according to the McDonald criteria (McDonald *et al*., 2001).

Statistical analysis

Mean (SD) atrophy rates in each subject group for each of the three methods were calculated. To investigate bias, a Bland-Altman plot of the BBSI and SIENA was generated, and differences in the mean and SD of each method were assessed within subject groups using paired *t*-tests and Pitman's test respectively. Where ANOVA suggested a significant difference in atrophy rates between subject groups ($p<0.05$), a two-tailed independent samples *t*-test was applied and the 95% CI of the difference in rate

was calculated. In addition, the ability of each method to distinguish between subject groups was assessed using logistic regression. Differences in the mean atrophy rate between CIS subjects who had developed MS at three year follow-up and those who had not was assessed using a two-tailed independent samples *t*-test and the 95% CI of the difference in mean rate was calculated.

6.2.3 Results

Mean (SD) brain atrophy rates within subject groups quantified by each technique are presented in Figure 6-1 and Table 6-2.

Figure 6-1 *Brain atrophy rates quantified using segmented brain volume difference, the BBSI and SIENA, in controls and subjects with CIS and RRMS. Subjects with CIS were divided into those who had not developed MS at three year follow-up (CIS (CIS)) and those who had (CIS (MS)). The two CIS subjects who did not have three year follow-up are included in the CIS (CIS) group (their atrophy rates were -0.58% year^{-1} and 0.28% year^{-1} using segmented brain volume difference, -0.41% year^{-1} and 0.00% year^{-1} using the BBSI and -0.16% year^{-1} and -0.22% year^{-1} using SIENA).*

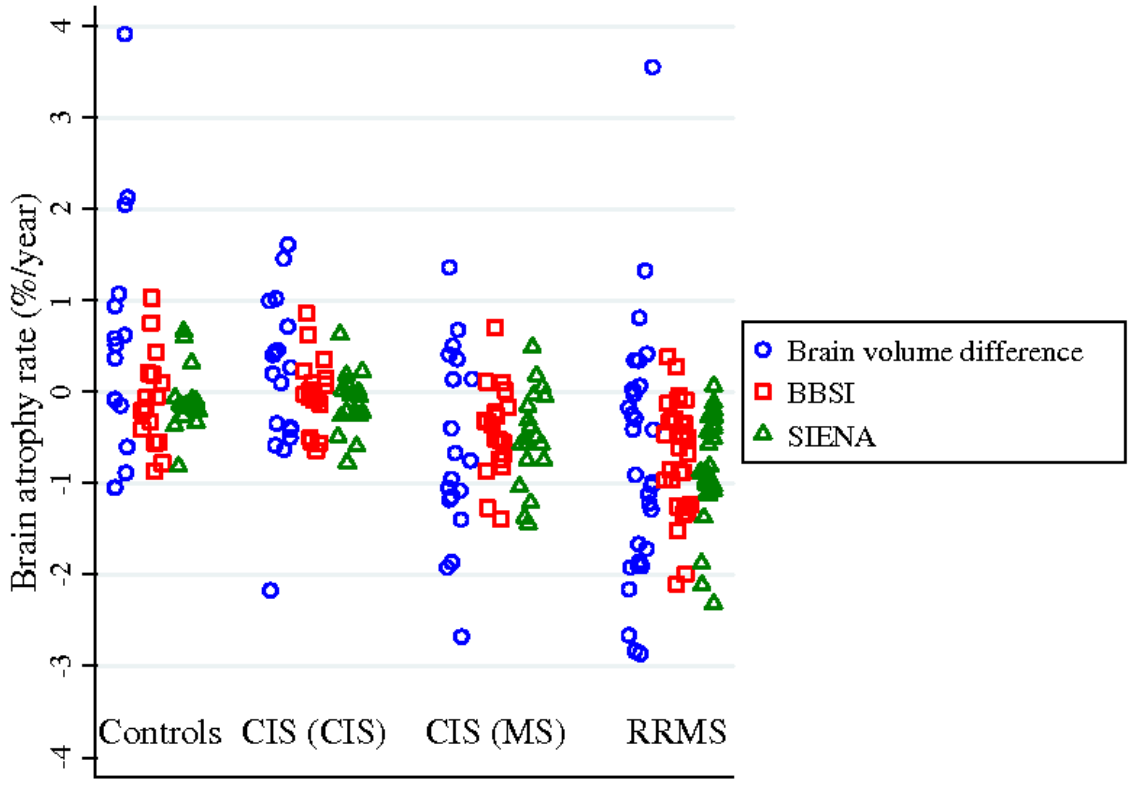


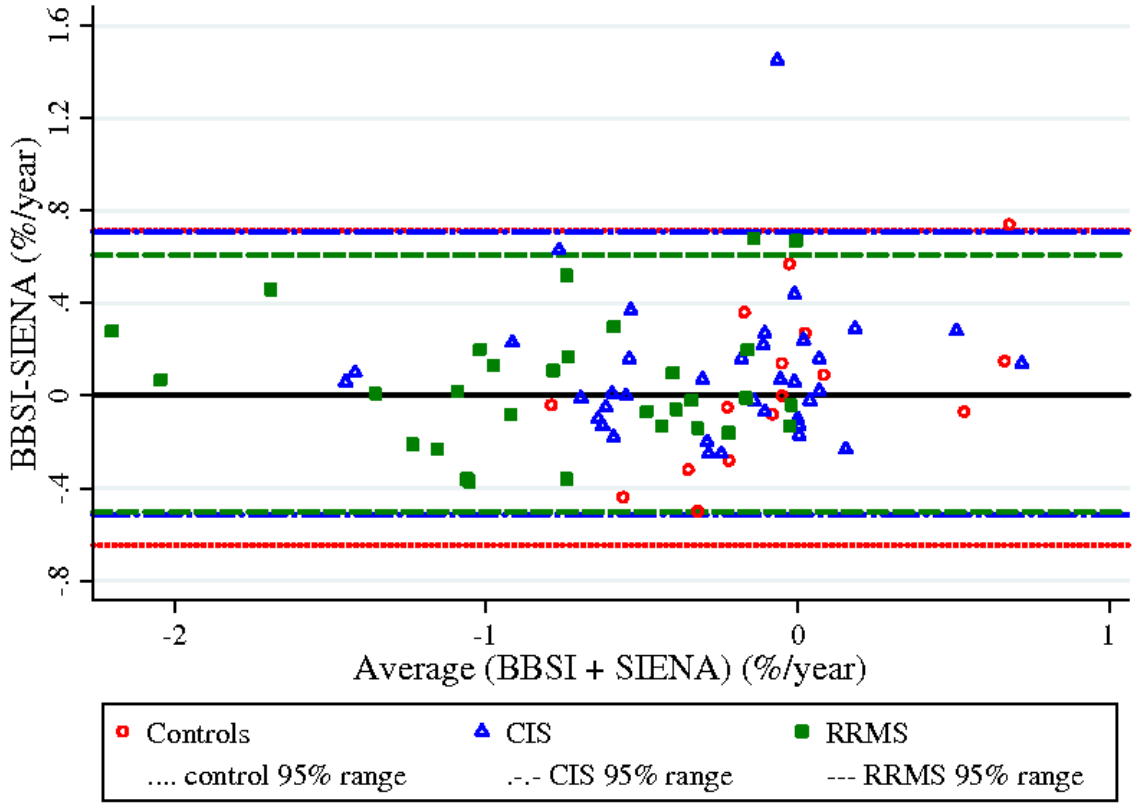
Table 6-2 *Mean brain atrophy rates quantified using segmented brain volume difference, the BBSI and SIENA, in controls and subjects with CIS and RRMS.*

	Mean/median (SD) atrophy rate (% year^{-1})		
	Controls	***CIS***	***RRMS***
BVD	0.51/0.44 (1.32)	-0.25/-0.28 (1.04)	-0.76/-1.02 (1.38)
BBSI	-0.04/-0.02 (0.52)	-0.22/-0.14 (0.48)	-0.72/-0.59 (0.60)
SIENA	-0.07/-0.10 (0.34)	-0.32/-0.22 (0.47)	-0.78/-0.78 (0.59)

The Bland-Altman plot of measures obtained by the BBSI and SIENA showed no significant bias between the two methods (Figure 6-2). No significant differences between mean atrophy rates quantified by BVD, BBSI and SIENA were observed in any of the subject groups. However Pitman's tests showed that the variance of atrophy rates within each subject group was significantly reduced when quantified either with the BBSI or SIENA compared with BVD (all $p<0.001$). In CIS and RRMS subject groups the variance in atrophy rates was very similar between the BBSI and SIENA, but was less similar in the control group ($p<0.05$). However, SIENA and the BBSI were closely correlated; the mean difference between the measures (BBSI minus SIENA) was 0.03% year^{-1} (SD 0.34, 95% CI -0.15 to 0.22, p=0.70) in controls, 0.10% year^{-1} (SD 0.31, 95% CI -0.007 to 0.20, p=0.07) in subjects with a CIS and 0.05% year^{-1} (SD 0.28, 95% CI -0.05 to 0.16, p=0.32) in subjects with RRMS.

There was no significant difference in mean age (p=0.28) between subject groups. Brain atrophy rates calculated using each technique increased between controls and patients with CIS and between patients with CIS and those with RRMS (Table 6-2). ANOVA indicated significant between group differences using each technique. The difference in mean brain atrophy rate between patients with CIS and control subjects using BVD was 0.76% year^{-1} (95% CI 0.08 to 1.44, p=0.0287). No significant difference was observed using the BBSI or SIENA. The difference in atrophy rates between patients with RRMS and controls was 1.27% year^{-1} (95% CI 0.42 to 2.11, p=0.0044) using BVD, 0.69% year^{-1} (95% CI 0.33 to 1.04, p=0.0003) using BBSI and 0.71% year^{-1} (95% CI 0.38 to 1.03, p=0.0001) using SIENA. Significant differences in brain atrophy rates between patients with RRMS and CIS were shown by the BBSI and SIENA: 0.50% year^{-1} (95% CI 0.24 to 0.76, p=0.0003) and 0.46% year^{-1} (95% CI 0.20 to 0.71, p=0.0007) respectively.

Figure 6-2 *Bland-Altman plot comparing brain atrophy rates quantified by SIENA and the BBSI, in controls and subjects with CIS and RRMS. The reference ranges are the values within which 95% of the differences between atrophy measurements from the BBSI and SIENA are expected to lie.*



Logistic regression demonstrated that subjects were 12 times (95% CI 2 to 60 times) more likely to be a RRMS subject than a control for each 1% year^{-1} increase in brain atrophy rate for the BBSI (p=0.0002) and 111 times (5 to 2600 times) more likely for SIENA (p<0.0001), compared with only twice as likely (1.2 to 4 times) with BVD (p=0.0039). SIENA was significantly better than both BVD (p=0.0002) and the BBSI (p=0.0058) at distinguishing patients with RRMS from control subjects.

Of the 35 patients with CIS who had three year follow-up, 19 had developed MS (17 RRMS, two SPMS) whilst 16 remained clinically isolated. Table 6-3 lists mean (SD) brain atrophy rates in both groups for each of the techniques. Mean brain atrophy rates were significantly different between patients who had developed MS and those who had not using each technique: 0.83% year^{-1} (95% CI 0.15 to 1.51, p=0.0187) using BVD, 0.45% year^{-1} (95% CI 0.14 to 0.76, p=0.0056) using the BBSI and 0.42% year^{-1} (95% CI

0.12 to 0.72, p=0.0078) using SIENA. In the 19 subjects who had developed MS, all methods showed that atrophy rates were significantly greater than in control subjects: 1.14% year^{-1} (95% CI 0.33 to 1.95, p=0.007) using BVD, 0.39% year^{-1} (95% CI 0.04 to 0.74, p=0.0282) using the BBSI, and 0.45% year^{-1} (95% CI 0.15 to 0.75, p=0.0048) using SIENA. No significant difference was observed between controls and CIS patients who did not develop MS.

Table 6-3 *Demographics and mean (SD) brain atrophy rates in subjects with CIS according to three year clinical status.*

	Subjects with MS at three years	***Subjects remaining CIS at three years***
Male:Female	4:15	8:8
Age, years (mean, SD)	35.6 (6.6)	33.3 (5.2)
EDSS at baseline (median, range)	1 (1-2)	1 (1-4)
BVD (% year^{-1})	-0.64 (1.04)	0.19 (0.93)
BBSI (% year^{-1})	-0.43 (0.50)	0.02 (0.38)
SIENA (% year^{-1})	-0.52 (0.50)	-0.10 (0.34)

6.2.4 Discussion

In this study brain atrophy rates quantified using BVD and two registration-based methods, the BBSI and SIENA, have been compared in controls and subjects with CIS and early RRMS. It has been shown that atrophy rates obtained using the BBSI and SIENA are well correlated and provide a higher level of measurement precision than volume subtraction based on manual segmentation methods. This study has also confirmed brain atrophy rates to be significantly greater in subjects with early RRMS than controls, whilst atrophy rates are already increased in some subjects presenting with CIS, particularly those who go on to develop MS. These results were not affected by the atrophy measurement technique used.

Although atrophy quantified from segmented BVD yielded similar mean rates to the registration-based techniques in patients with CIS and RRMS, there is a much higher variance associated with this measure. In agreement with these findings, a study investigating a segmentation-based technique and SIENA found that SIENA reduced the

standard deviation of longitudinal atrophy measurements by over half those obtained using the segmentation technique (Sormani *et al*., 2004). Another study investigating brain atrophy in PPMS using serial segmentation and SIENA also demonstrated that the variance of measurements was lower using SIENA (Sastre-Garriga *et al*., 2005a). As already described, measurement precision is particularly important if brain atrophy is to be used as a primary outcome measure in trials of putative disease-modifying drugs. For a given number of subjects the power to detect a significant difference between groups is driven by the standard deviation of the measurement technique. Sormani *et al.* found that the power to detect a difference in atrophy rates between the placebo and treatment arms of a trial was 32% using the segmentation-based technique but 73% using SIENA, due to the difference in standard deviation between methods (Sormani *et al*., 2004). By increasing measurement precision and statistical power, sample sizes can be reduced which in turn reduces the length and cost of clinical trials (Fox *et al*., 2000a). This study suggests that the BBSI and SIENA have almost equal power. In addition to the BBSI and SIENA providing direct quantification of atrophy through image subtraction, the registration process also compensates for variations in the MR scanner over time, which can lead to changes in apparent voxel sizes. It is a possibility that the greater variance in atrophy measures that was observed using the BVD in this study is due not only to segmentation errors, but that the volumes were not normalised. Some segmentation methods, such as the BPF do perform a normalisation step (Rudick *et al*., 1999), but may still be subject to errors in segmentation at serial time-points.

Results from multiple logistic regression suggest that SIENA may be slightly more sensitive than the BBSI in distinguishing subjects with RRMS from controls, although both methods provided greater sensitivity than BVD. This is confirmation of the importance of using a precise measurement technique. Similarly to these findings, a previous study which compared methods derived from SIENA and the BBSI found that the SIENA-derived method gave better group separation between control subjects and patients with Alzheimer's disease compared with the BBSI-derived method (Gunter *et al*., 2003). However in the work presented in this chapter, the greater sensitivity of SIENA is most likely due to the lower variance seen in the control group using this technique. No difference between SIENA and the BBSI was observed in distinguishing subjects with CIS from controls, and although segmented BVD showed a significantly greater brain atrophy rate in CIS than control subjects, this was probably due to the large variance of

measurements in all subject groups using segmentation, and the outliers seen in the control group leading to an unexpected positive atrophy rate (brain "growth").

Recently, a cross-validation study of the BBSI and SIENA has been performed in a cohort consisting of 23 control subjects and 45 patients with Alzheimer's disease (Smith *et al*., 2007), with MRI data available at seven time-points enabling the quantification of atrophy over multiple intervals. This study confirmed the agreement shown in the present investigation between the BBSI and SIENA, with a median absolute difference in atrophy of 0.25%, which compares to a mean difference of 0.07% year^{-1} found in this study. Whilst there was a trend for greater brain atrophy rates to be detected using SIENA than the BBSI in the study presented here, Smith *et al.* found that SIENA gave a 20% larger estimate of atrophy than the BBSI. This difference in estimates of atrophy is to be expected from the original report of the BBSI (Freeborough & Fox, 1997) where it was reported that the BBSI in effect scaled atrophy by about 0.8. Estimates of effect size (Cohen's d) by Smith *et al.* showed the BBSI to be slightly more powerful.

Considering whether the results from the BBSI and SIENA presented in this study are typical of those that might be expected in such a cohort, the atrophy rates observed in subjects with RRMS appear to be similar to those found in previous investigations. One study which applied the BBSI found an annual median brain atrophy rate of -0.8% year^{-1} (Fox *et al*., 2000b). Likewise, application of SIENA to 34 subjects with RRMS showed a median brain atrophy rate of -0.7% year^{-1} (SD 0.9) (Rovaris *et al*., 2003). A larger analysis of 105 subjects using SIENA found mean brain atrophy was -0.9% (SD 1.2) over only nine months (Sormani *et al*., 2004).

Although no previously published studies have used the BBSI to quantify brain atrophy in subjects with CIS, several studies have applied SIENA to these subjects. In agreement with these findings, one study of 20 subjects presenting with a CIS found a median brain atrophy rate of -0.3% year^{-1} (SD 0.6) (Rovaris *et al*., 2003). However a larger study of 38 subjects, who had not developed clinically definite MS by an 18 month follow-up, found a median brain atrophy rate of -0.58% year^{-1} (IQR -1.02 to -0.24) (Paolillo *et al*., 2004), whilst investigation of beta interferon treatment in subjects with CIS found that brain atrophy rate in around 100 placebo subjects was -0.83% year^{-1} (SD 1.09) during the first year of study and -0.67% year^{-1} (SD 1.10) during the second year of study (Filippi *et al*.,

2004). Within any group of subjects presenting with a CIS, some may develop MS whilst others remain clinically isolated, and the heterogeneity of CIS subject groups within different studies may account to some extent for the different atrophy rates observed. When the CIS subjects were divided into those that had developed MS at three year follow-up and those who had not, all techniques showed that the atrophy rate in the MS group approached that seen in subjects with established RRMS, whereas in subjects remaining clinically isolated the atrophy rate approached that of controls. Other investigations have shown rates of ventricular enlargement (Brex *et al*., 2000) and GM atrophy (Dalton *et al*., 2004) to be greater in CIS subjects developing MS than in those who do not. In a trial of beta interferon in CIS, brain atrophy was used as a secondary outcome measure, and atrophy rates of -0.92% were observed during the first year of study in subjects who had developed MS at two years, compared with -0.56% in subjects remaining stable (Filippi *et al*., 2004). Over the second year of the trial atrophy was only -0.64% in subjects who had developed MS compared with -0.50% in subjects remaining stable. These rates are greater than those observed in this study, particularly for subjects remaining CIS, but this may be the result of the treatment in some subjects (with a treatment-associated reduction in oedema), the shorter follow-up period of two years within which MS could be diagnosed, or the selection of only CIS patients with MRI abnormalities (some of the CIS patients in this study had a normal scan). One of the limiting factors when investigating the differences between subjects with CIS who develop MS and those who do not is the length of follow-up. It is possible that subjects who remained clinically isolated at follow-up may still develop MS, and this may explain the difference in atrophy rates observed between other studies and the present one.

In summary, registration-based techniques applied to three-dimensional MRI acquisitions provide more precise measurement of brain atrophy rates than segmentation-based techniques, with the BBSI and SIENA providing comparable results. Accurate evaluation of brain atrophy is necessary if it is to be used for monitoring treatment efficacy in controlled trials. Although not investigated in this study directly, accurate evaluation may also aid in the identification of patients most likely to obtain long-term benefit from disease-modifying treatment.

6.3 Comparison of the BBSI and SIENA with other cerebral atrophy measurement techniques, and reliability of measurements

6.3.1 Introduction

The previous chapters have confirmed that brain atrophy can be successfully measured at the earliest stages of MS using two registration-based techniques. In addition it has been shown that these measures may be more precise markers of progression than atrophy rates derived from the subtraction of absolute brain volumes. Other techniques based on different methodology or regional measurements have been proposed as markers of whole brain atrophy however (Chard *et al*., 2002b; Fox *et al*., 2000b; Losseff *et al*., 1996; Schott *et al*., 2005), which may provide a similar sensitivity and precision of measurement to the BBSI and SIENA.

One such measure is ventricular enlargement, and previous studies in MS and CIS have shown ventricular enlargement at the earliest stages of the disease (Brex *et al*., 2000; Dalton *et al*., 2002a; Kalkers *et al*., 2002). Furthermore, in some studies significant ventricular enlargement has been observed when no significant change in WBV has been detected (Horakova *et al*., 2007; Turner *et al*., 2003), suggesting that this measure may be more sensitive to small changes, as tissue loss throughout the whole brain may result in relatively large increases in CSF spaces. Manual outlining of the lateral ventricles can be performed quickly and with a high degree of accuracy given the high contrast brain-CSF boundary. In the last few years the automated ventricular boundary shift integral (VBSI) technique has been applied, which directly quantifies ventricular enlargement following the accurate registration of local ventricular regions, using the same methodology as the BBSI (Schott *et al*., 2005). It has been shown to provide highly similar measures to volume subtraction following segmentation.

The CCV is defined as a region that includes a large proportion of the lateral ventricles (Losseff *et al*., 1996), and measures of this ROI may therefore be similarly sensitive to global tissue loss as ventricular enlargement. One of the main advantages of this measure is that segmentation of the superior and inferior limits of the brain is avoided which, firstly, may decrease the time required for analysis relative to whole brain segmentation methods and, secondly, could reduce segmentation errors that may occur due to the complex folding of gyri and sulci or due to image artefacts away from the centre of the

image. Significant decreases in CCV have been shown in patients with MS (Losseff *et al*., 1996; Rovaris *et al*., 2000).

As mentioned in Chapter 6.2.4 it is possible that the greater variance in atrophy measures observed following the subtraction of brain volumes was due not only to segmentation errors, but that the volumes were not normalised. SPM software allows normalised measures of brain, GM and WM to be generated. Although MS lesions occur within the GM they are thought to be associated with less inflammation (Bø *et al*., 2003a) and therefore measures of GM atrophy could be more sensitive markers of neuroaxonal loss and disease progression. Studies have shown progressive GM atrophy occurring in patients presenting with CIS and RRMS (Chard *et al*., 2004; Dalton *et al*., 2004). Moreover it has been shown to progress at a greater rate than atrophy of the WM (Dalton *et al*., 2004; Tiberio *et al*., 2005), and be correlated with disability (Sanfilipo *et al*., 2005). Atrophy of the cortex has also been shown to correlate with measures of cognitive impairment in people with MS (Portaccio *et al*., 2006), and therefore measures may be of particular clinical relevance.

As the methodologies behind these other techniques differ from those presented in Chapter 6.2, it is important to determine the relative power of these measures to detect brain atrophy and, by implication, disease progression in MS. The primary aim of this study was to compare these techniques directly with the BBSI and SIENA. The relative sensitivity, precision and effect sizes are investigated for BBSI, SIENA, ventricular enlargement, VBSI, CCV, SPM BPF, SPM GMF and SPM WMF. As no standard software has been adopted for the measurement of CCV, a secondary aim of this study was to investigate the relative advantages of two different software packages (Excalp and MIDAS) that could be used to obtain this volume. In addition, one of the potential problems with longitudinal CCV measurement is ensuring that subjects are positioned consistently within the scanner, so that serial measurements are obtained on the same ROI. Therefore it was also investigated whether registration of images could improve atrophy rates estimated from CCV.

One of the important features of any potential outcome measure for a clinical trial is its reliability (the reproducibility of a measurement when repeated in the same subject) (Lachin, 2004). Random measurement error will reduce reliability and decrease the power

to detect a treatment effect. Whilst reliability can be calculated by performing repeated measurements on the same scan, this does not take into account the degree to which the measurement is influenced by scanner-related variability. For example, an inaccurate but fully automated measure would produce exactly the same result when applied twice to the same scan. If, however, measurements on two scans acquired on the same day on the same subject produce very different values, the measure (scan plus analysis method) must be deemed to lack reproducibility. Therefore this study also analyses the relative reliability and consistency of different measurement techniques based on same-day scan-rescan MRI.

6.3.2 Methods

Subjects and MR imaging

Fourteen controls (five male, mean age 35.3 years (SD 6.3)), 41 subjects presenting with a CIS (15 male, mean age 34.3 years (SD 6.4)) and 29 subjects with RRMS (eight male, mean age 37.3 years (SD 7.4)) were identified with both T1-weighted coronal volumetric FSPGR and T1-weighted 2D CSE MR imaging. Details of the acquisitions are described in Chapter 3.3 and the FSPGR sequence was the same for the three subject groups. In controls and patients with RRMS CSE images had 5mm thick slices. In subjects presenting with CIS, CSE images had 3mm thick slices and were acquired following the administration of 0.1mmol/kg Gd-DTPA. MRI was acquired at baseline and approximately one year later in all patients, with a mean interval of 1.17 years (SD 0.24).

Twenty-two older healthy control subjects were identified from a longitudinal project undertaken at the Institute of Neurology, aimed at determining the shortest interval required to detect volumetric change based on MRI, to distinguish AD subjects from normal controls (11 male, mean age 69.6 years (SD 7.3)). Subjects had two MRI scans acquired on the same day (without being removed from the scanner) and a repeat scan approximately one year later (mean interval 1.0 years (SD 0.0)), which were used to perform an assessment of the reliability of measurement techniques. Coronal T1-weighted MRI was acquired in these subjects according to the protocol described in Chapter 3.3.3.

MRI analysis

All methods were applied to FSPGR MRI except for estimation of CCV which was determined on CSE images.

(i) BBSI and SIENA

The BBSI and SIENA were obtained as described in Chapter 6.2.2.

(ii) Segmented ventricle volume difference (VVD)

Images were placed into standard space based on the MNI-152 brain image (www.bic.mni.mcgill.ca/cgi/icbm_view) (Mazziotta *et al*., 2001) and using a 9dof6 registration (the transformation matrix is determined using 9dof, but the image is transformed using only translations and rotations). Each repeat brain in standard space was subsequently registered to the corresponding baseline brain in standard space using affine registration. An upper threshold value representing 60% of the mean brain intensity was used to delineate ventricles (which included the lateral ventricles and temporal horn) on registered images using MIDAS (Freeborough *et al*., 1997). This threshold excluded brain, whilst a lower threshold set at zero was used to include CSF. Baseline ventricle volume was subtracted from repeat ventricle volume and corrected for scan interval.

(iii) Ventricular boundary shift integral (VBSI)

Using the ventricle regions obtained in (ii), a local 6dof registration was performed to positionally match ventricular regions between the baseline and registered repeat images in standard space. The VBSI was calculated over this local region in the same manner as the BBSI. Calculation of the VBSI was also performed using only the baseline ventricle ROI. VBSI measurements were corrected for scan interval.

(iv) Central cerebral volume

The "starting slice" (the most inferior slice of the ROI) was determined on each scan by identifying the velum interpositum cerebri, guidelines for which were developed in association with an experienced neuroradiologist (Appendix 3). For repeat images, the choice of starting slice on the baseline image was referred to in order to match the starting slices as closely as possible. Excalp and MIDAS were used to obtain an ROI containing this starting slice and the three (RRMS patients) or five (CIS patients) slices above it. Prior to application of Excalp (see Chapter 3.4.5) each image slice to be processed was saved as a separate file. Excalp was used to automatically strip the skull from each slice and the morphological opening operator used to separate the brain from other components in the image was set to a diameter of 10mm. Any remaining areas of non-brain were removed by manual editing in DispImage (see Chapter 3.4.4). An in-house script was

used to calculate the volume of each slice by summing the number of voxels multiplied by the voxel dimensions. CCV was the sum of the slice volumes. Baseline CCV was subtracted from the repeat CCV and the difference expressed as a percentage of baseline CCV volume, before correcting for scan interval to produce a rate of change.

The method used to obtain CCV using MIDAS was similar to that used to obtain WBV. Intensity thresholds were set to exclude voxels brighter e.g. dura/scalp, and darker e.g. CSF, than these values over the whole image. In contrast to whole brain segmentation the most inferior slice was set to be the "starting slice", which therefore excluded voxels inferior to this slice. Erosion and a conditional dilation of the ROI was then performed. Following this the most superior slice of the CCV was selected and the voxels on all slices above this were deleted. Manual editing was performed on the resulting region if required and the total CCV was calculated automatically within MIDAS. As before, baseline CCV was subtracted from repeat CCV and expressed as a percentage of baseline volume before correcting values for scan interval.

Registration of CSE images was performed by registering the MNI-152 standard space reference scan (www.bic.mni.mcgill.ca/cgi/icbm_view) (Mazziotta *et al*., 2001) to each image, using the FLIRT registration package (Chapter 3.4.2) (12dof) and maximising the normalised mutual information (Studholme *et al*., 1999). The brain segmentation available in FSL for this reference scan was transformed to each individual scan by applying the corresponding transformation parameters obtained from image registration, thereby creating an approximate brain mask on each scan in its native space. Using the approximate brain regions, each baseline image was subsequently registered to the MNI-152 standard space brain template using a rigid 9dof6 registration with renormalised sinc interpolation (Thacker *et al*., 1999). Transformation parameters were subsequently applied to the approximate baseline brain region. Repeat images were registered to standard space baseline images using 12dof and renormalised sinc interpolation. All resulting registrations were checked for accuracy. Starting slices were determined for each subject and CCV was obtained on each scan using MIDAS as described above.

(v) SPM fractional measures

Images were reoriented axially, maintaining all voxel dimensions. Using SPM99 software (Ashburner & Friston, 1997) with inhomogeneity correction (Chard *et al*., 2002c), GM,

WM and CSF segments were automatically generated. Lesions were contoured on each of the images using the semi-automated DispImage software (Chapter 3.4.4) to create a lesion mask which was used to generate the final binary masks of GM, WM, CSF and lesions in SPM. Tissue inferior to the base of the cerebellum was excluded from the masks and all segmentations were visually assessed. Tissue volumes were subsequently determined from each binary mask and BPF, GMF and WMF were calculated as follows:

$$BPF = \frac{\text{grey matter} + \text{white matter} + \text{lesion volume}}{\text{total volume within surface contour (brain tissue volume} + \text{CSF)}}$$

$$GMF = \frac{\text{grey matter volume}}{\text{total volume within surface contour (brain tissue volume} + \text{CSF)}}$$

$$WMF = \frac{\text{white matter volume} + \text{lesion volume}}{\text{total volume within surface contour (brain tissue volume} + \text{CSF)}}$$

Changes in tissue fractions were obtained by subtracting the baseline from repeat estimates, and expressed as a percentage of baseline values.

Reliability

No scan-rescan CSE images were available and therefore the reliability of CCV measures was not assessed. Cross-sectional volume measurements were made on all images from the 22 older control subjects (ventricle volume, BPF, GMF and WMF). Atrophy rate was subsequently determined between scan pairs for these methods and for the BBSI, SIENA and the VBSI: i) baseline A to baseline B, ii) baseline A to one year repeat, iii) baseline B to one year repeat.

Statistical analysis

Mean (SD) atrophy rates for each of the methods was calculated for each subject group, and the effect size (μ/σ) was calculated for patients with RRMS. As ventricular enlargement was not expressed as a percentage of baseline, due to the large normal variation that can occur in baseline ventricular size, these results were analysed separately. In each subject group a joint Wald test was used as an overall test of differences in mean percentage atrophy rates. Where a significant difference was

observed in RRMS subjects ($p<0.05$), paired *t*-tests were used to investigate pairwise differences. Likewise, a generalisation of Pitman's test to more than two observations was performed within each subject group to determine if there was a significant difference in the variance of atrophy rates using the different methods (Han, 1969). Where a significant difference was observed in RRMS subjects ($p<0.05$), Pitman's test was applied pairwise.

Measures of VVD were correlated with whole brain atrophy rates from SIENA using Pearson's correlation coefficient. Bland-Altman plots of VVD and the VBSI were generated to assess visually whether there was any bias in measures, and a joint Wald test was applied to the three measures of ventricular enlargement to determine if there was a significant ($p<0.05$) difference in mean rates. Likewise, generalisation of Pitman's test was used as overall investigation of differences in the variance of measures obtained using the three techniques. Where significant differences were observed in RRMS subjects ($p<0.05$), pairwise *t*-tests and Pitman's tests were applied to observations.

Assessing reliability, the mean value of the scan-rescan volume change estimate was calculated. This should be zero if there is no systematic bias in the estimation of atrophy. A one-sample two-tailed *t*-test was used to investigate whether the differences observed were significantly different from zero. Assessing the consistency of annual atrophy rates that were calculated from each of the two baseline images, Bland-Altman plots were generated to assess whether there was any bias in measurements. In addition, a paired *t*-test was performed and the mean difference (SD, 95% CI) between the two measurements was calculated for each method. The coefficient of reliability (intraclass correlation coefficient (ICC)) between the two longitudinal atrophy measures was calculated for each method (Lachin, 2004). This value gives an estimate of the proportion of variation that is not due to measurement error. Significant differences between ICCs were investigated by calculating the 95% bootstrap CI for the difference between the ICC of two techniques.

6.3.3 Results

Excalp failed on one control subject and SPM failed on one patient with RRMS, therefore these two subjects were excluded from further analysis. Both Excalp and MIDAS CCV methods required minor editing of regions following initial processing. All tissue segmentations from SPM were acceptable on visual inspection. Mean (SD) rates of

atrophy and ventricular enlargement, plus effect sizes for each method, are given in Table 6-4, whilst Figure 6-3 and Figure 6-4 show the changes measured on individual subjects.

Table 6-4 *Mean rates of brain atrophy and ventricular enlargement quantified from eleven different measurement techniques, in controls and subjects with CIS and RRMS. Effect sizes (μ/σ) are given for subjects with RRMS.*

		Controls	***CIS***	***RRMS***	***Effect size***
Mean (SD) atrophy rate (% year^{-1})	*Excalp CCV*	0.69 (1.02)	-0.48 (1.05)	0.66 (2.93)	0.20
	MIDAS CCV	0.96 (1.40)	-0.21 (1.29)	-0.18 (1.53)	0.15
	MIDAS CCV (registered images)	0.40 (0.34)	-0.09 (0.46)	-0.26 (0.48)	0.53
	BBSI	0.04 (0.51)	-0.20 (0.49)	-0.70 (0.55)	1.26
	SIENA	0.00 (0.31)	-0.30 (0.48)	-0.75 (0.51)	1.38
	BPF (SPM)	-0.41 (1.16)	-0.59 (1.39)	-1.16 (1.26)	0.92
	GMF (SPM)	-0.33 (1.22)	-0.97 (1.87)	-1.35 (1.57)	0.86
	WMF (SPM)	-0.55 (1.39)	0.23 (1.30)	-0.76 (1.65)	0.46
Mean (SD) ventricular enlargement (ml year^{-1})	*VVD*	-0.23 (0.68)	0.41 (1.07)	1.32 (1.15)	0.99
	VBSI (both ROIs)	-0.20 (0.67)	0.34 (0.94)	1.22 (1.03)	1.02
	VBSI (baseline ROI)	-0.17 (0.68)	0.31 (0.86)	1.16 (1.00)	0.99

CCV, central cerebral volume; BPF, brain parenchymal fraction; SPM, statistical parametric mapping; GMF, grey matter fraction; WMF, white matter fraction; VVD, segmented ventricular volume difference; VBSI, ventricular boundary shift integral; ROI, region of interest.

Figure 6-3 *Atrophy rates quantified from eight different measurement techniques in controls and subjects with CIS and RRMS.*

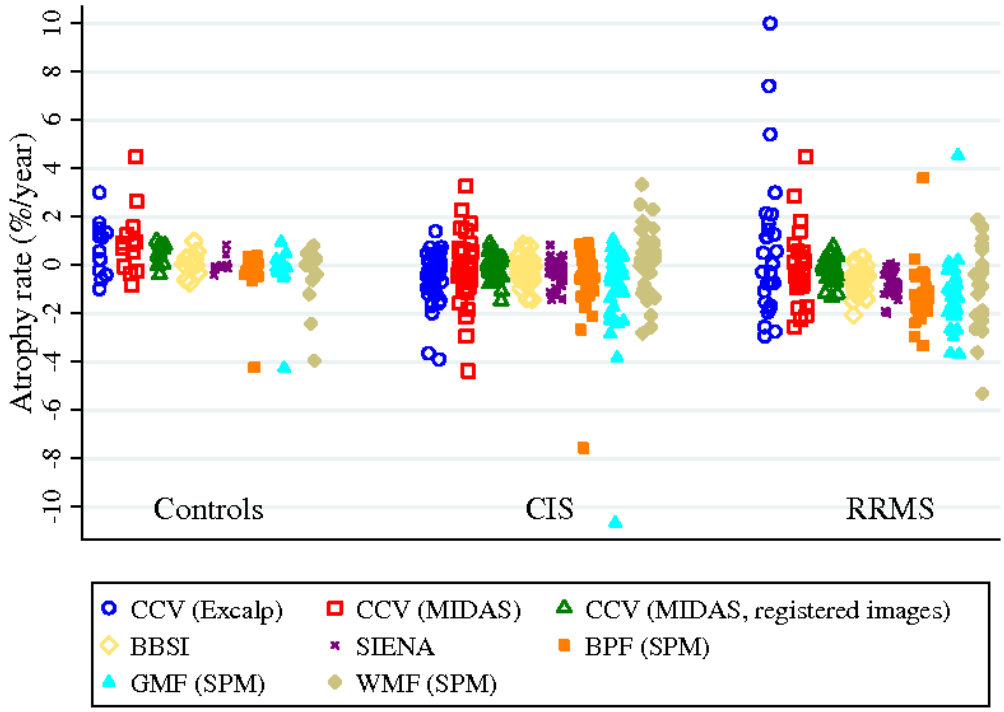
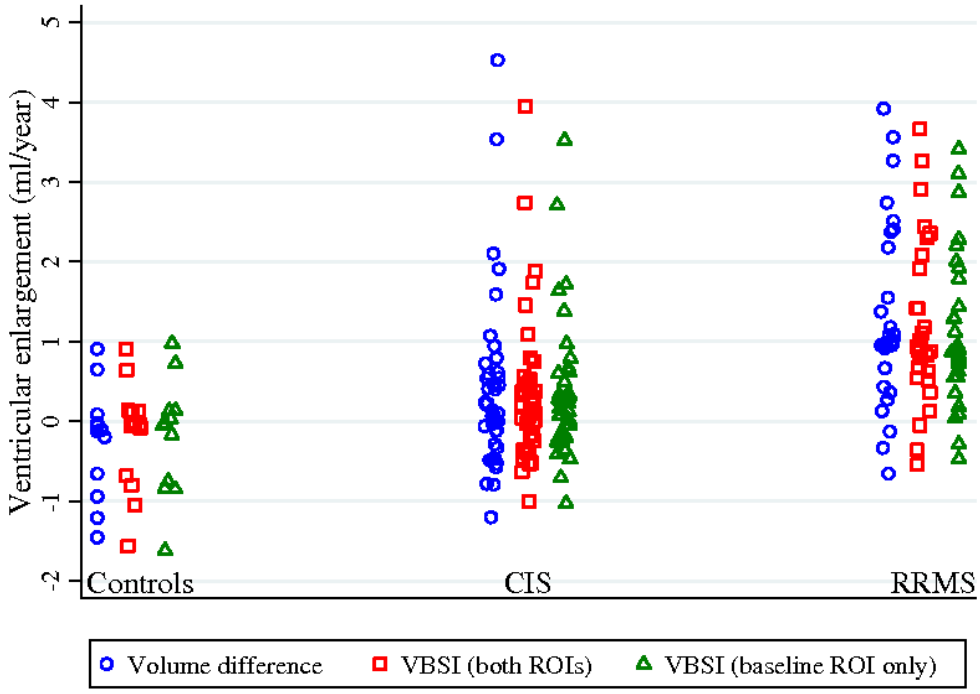


Figure 6-4 *Rates of ventricular enlargement quantified from segmented ventricular volume difference and the VBSI, in controls and subjects with CIS and RRMS.*



CCV, central cerebral volume; BPF, brain parenchymal fraction; SPM, statistical parametric mapping; GMF, grey matter fraction; WMF, white matter fraction; VBSI, ventricular boundary shift integral; ROI, region of interest.

Ventricular enlargement and atrophy rates calculated using BPF, GMF, and CCV on registered images, increased between controls and patients with CIS, and between patients with CIS and those with RRMS. Change in WMF, and CCV measured on unregistered images, were less consistent. Positive atrophy rates (brain "growth") were seen in control subjects using CCV methods, and in RRMS subjects when using the Excalp CCV method. Small effect sizes were also observed when atrophy was quantified using CCV, although registration of images did improve statistical power slightly. The BBSI and SIENA gave the largest effect sizes, with all three measures of ventricular enlargement giving relatively high statistical power also. BPF and GMF gave slightly smaller effect sizes, whilst those obtained using atrophy of the WMF were poor.

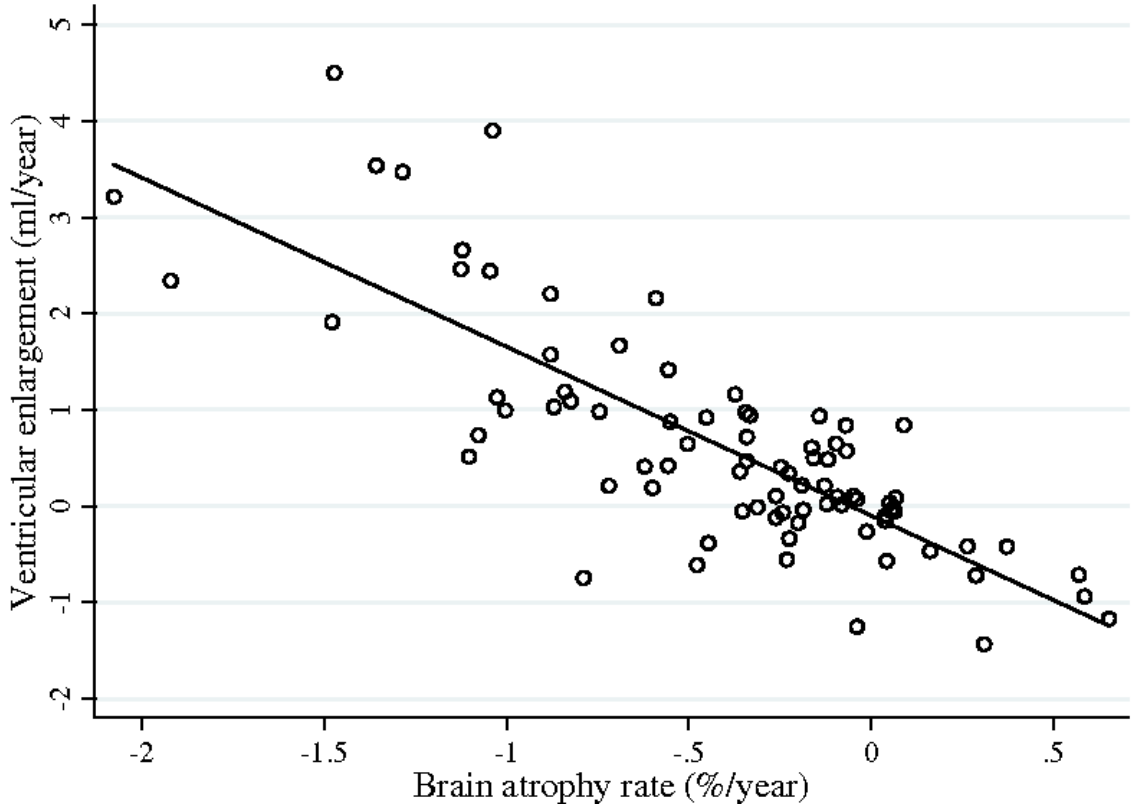
There was evidence of a significant difference in mean percentage atrophy rates given by the different methods in each subject group ($p=0.0028$ for controls, $p=0.0003$ for CIS and $p<0.0001$ for RRMS). Paired *t*-tests in RRMS subjects showed that mean rates of GMF atrophy were significantly greater (more negative) than rates quantified by the BBSI ($p=0.0408$), SIENA ($p=0.0405$), and CCV (Excalp $p=0.0016$, MIDAS $p=0.0128$, registered images $p=0.0011$). Rate of BPF atrophy was significantly greater than rate of atrophy from CCV (Excalp $p=0.0058$, MIDAS $p=0.0175$, registered images $p=0.0008$), whilst the BBSI and SIENA gave significantly greater atrophy rates than CCV measured using Excalp (BBSI $p=0.0213$, SIENA $p=0.0169$) or MIDAS on registered images (BBSI $p=0.0001$, SIENA $p<0.0001$).

There was also evidence of a significant difference in the variance of atrophy rates between techniques ($p=0.0032$ for controls, $p<0.0001$ for CIS and $p=0.0012$ for RRMS). Pairwise Pitman's tests revealed that the variance of atrophy rates was significantly reduced when quantified with the BBSI and SIENA compared with SPM fractional atrophy rates, and CCV atrophy rates measured on unregistered images (all $p<0.001$). Percentage rate of change of CCV measured on registered images was significantly less variable than SPM fractional atrophy rates and measures of CCV from unregistered images. BPF measurements were significantly more variable than measurements of GMF ($p=0.016$).

Ventricular enlargement was shown to correlate well with whole brain measures of atrophy by SIENA (Figure 6-5). The correlation of SIENA with VVD, VBSI with both

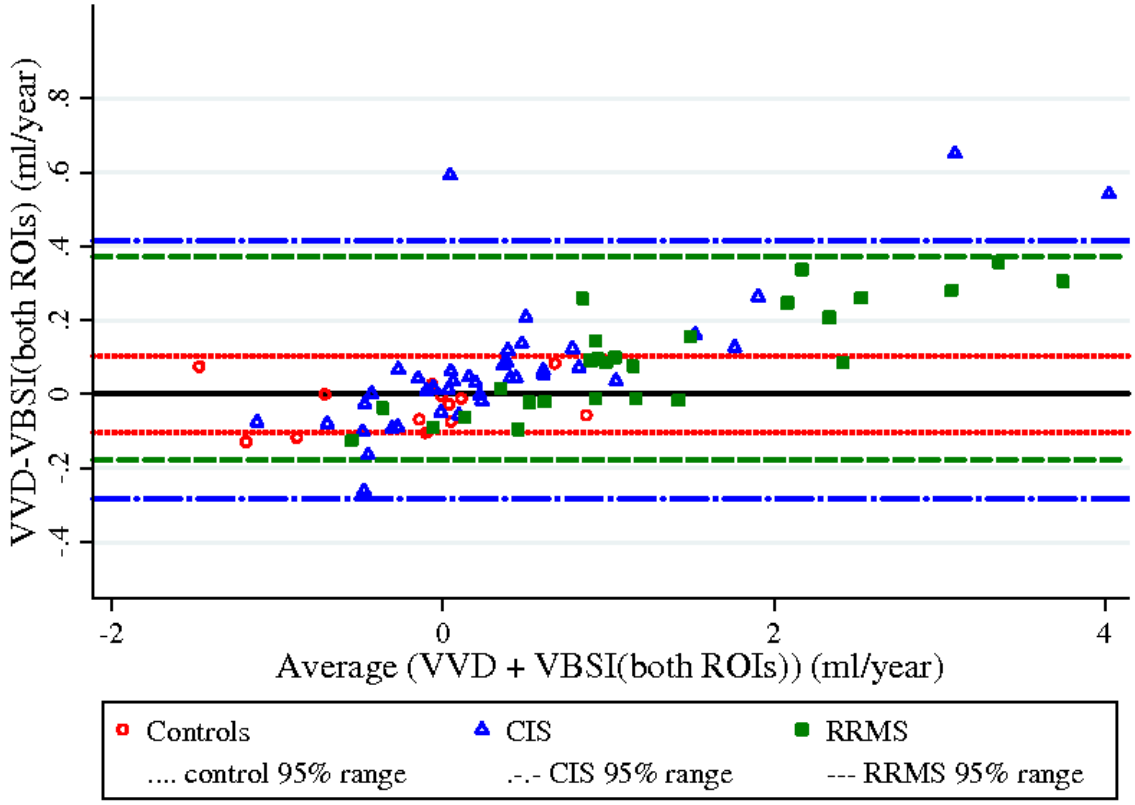
ROIs and VBSI with the baseline ROI only was r=-0.80, -0.80 and -0.79 respectively (all p<0.001).

Figure 6-5 *Correlation between rate of ventricular enlargement (segmented ventricle volume difference) and rate of whole brain atrophy (SIENA).*



Pairwise Bland-Altman plots of the different measures of ventricular enlargement showed that as ventricular enlargement increased there was a trend for the VBSI (both methods) to underestimate the volume change compared with the VVD (Figure 6-6). This was confirmed by the joint Wald test, where in subjects with RRMS there was evidence of a significant difference in rates of ventricular enlargement (p=0.0009). In these subjects the mean difference between VVD and the VBSI using both ROIs (VVD-VBSI) was 0.10ml $year^{-1}$ (SD 0.14, 95% CI 0.04 to 0.15, p=0.0011). Similarly the mean difference between VVD and the VBSI using only the baseline ROI was 0.16ml $year^{-1}$ (SD 0.24, 95% CI 0.07 to 0.25, p=0.0015). Looking at the two VBSI measurements, mean difference (VBSI (both ROIs) – VBSI (baseline ROI only)) was 0.06ml $year^{-1}$ (SD 0.14, 95% CI 0.006 to 0.11, p=0.0306).

Figure 6-6 *Bland-Altman plot comparing ventricular enlargement measures quantified from the difference in segmented ventricular volume and the ventricular boundary shift integral. The reference ranges are the values within which 95% of the differences between measurements from the two methods are expected to lie. VVD, segmented ventricular volume difference; VBSI, ventricular boundary shift integral (which was calculated using both baseline and repeat ventricular regions of interest (ROI)).*



In the CIS and RRMS subject groups there was evidence of a significant difference in the variance of rates of ventricular enlargement obtained using the three techniques ($p<0.0001$). Pairwise analysis revealed that the variance of rates was significantly greater when quantified by VVD compared with the VBSI (both $p<0.001$). As the relative changes in mean and variance of measures was similar, the effect sizes were almost identical for the three measures.

Reliability and consistency of brain atrophy measures

The mean difference in volume between scan-rescan baseline images is given in Table 6-5 for each method. There was evidence that the difference was significantly different from zero when quantified using the BBSI.

Table 6-5 *Scan-rescan analysis using the different atrophy measurement techniques in 22 control subjects. Mean volume difference between the two images and the results of a one-sample t-test are given.*

	Volume change*[a]** ***(mean (SD))	***One-sample t-test*** ***(p-value)***
BBSI	0.099% (0.21)	0.04
SIENA	0.016% (0.32)	0.81
BPF (SPM)	0.083% (0.83)	0.65
GMF (SPM)	0.090% (1.36)	0.76
WMF (SPM)	0.072% (1.00)	0.79
VVD	-0.043ml (0.27)	0.47
VBSI (both ROIs)	-0.091ml (0.35)	0.23
VBSI (baseline ROI only)	-0.185ml (0.45)	0.07

[a]*expressed as a percentage of volume 'A' except for ventricular measurements. BPF, brain parenchymal fraction; SPM, statistical parametric mapping; GMF, grey matter fraction; WMF, white matter fraction; VVD, segmented ventricle volume difference; VBSI, ventricular boundary shift integral; ROI, region of interest.*

Looking at the consistency of atrophy rates quantified from the two baseline images to repeat one year image, no bias was evident from inspection of the Bland-Altman plots (not shown). This was confirmed by paired *t*-tests, and mean differences calculated between the two measures were less than 0.06% for percentage atrophy rates and less than 0.2ml for ventricular enlargement (Table 6-6). The ICC between the atrophy rates calculated on repeated one year intervals was highest for VVD and was significantly greater than those for all other measures except the BBSI. The ICCs of the BBSI and VBSI using both ventricular regions were significantly larger than those of SIENA and the VBSI using only the baseline region. The reliability of SPM-derived fractional volumes was significantly lower than those of other measures.

Table 6-6 *Consistency of different atrophy measurement techniques. In 22 control subjects using each of the measurement techniques, brain atrophy or ventricular enlargement was quantified twice between baseline and one year, for each of the two (scan-rescan) baseline images. For each technique the mean difference in rates calculated from the two baseline images, and results of a paired t-test on all results for a technique are given. The intraclass correlation coefficient (ICC) of repeated measures is also given for each technique.*

	Difference in atrophy rate[a] (mean (SD) (95% CI))	***Paired t-test (p-value)***	***ICC***
BBSI	0.047% year^{-1} (0.226, -0.148 to 0.053)	0.34	0.91
SIENA	0.033% year^{-1} (0.312, -0.109 to 0.175)	0.63	0.86
BPF (SPM)	0.044% year^{-1} (0.839, -0.337 to 0.426)	0.81	0.65
GMF (SPM)	0.031% year^{-1} (1.371, -0.593 to 0.655)	0.92	0.64
WMF (SPM)	0.058% year^{-1} (1.016, -0.404 to 0.521)	0.80	0.76
VVD	-0.048ml year^{-1} (0.274, -0.173 to 0.077)	0.43	0.94
VBSI (both ROIs)	-0.109ml year^{-1} (0.345, -0.266 to 0.048)	0.16	0.91
VBSI (baseline ROI only)	-0.199ml year^{-1} (0.455, -0.406 to 0.008)	0.06	0.86

[a]*expressed as a percentage of baseline volume 'A' except for ventricular measurements, BPF, brain parenchymal fraction; SPM, statistical parametric mapping; GMF, grey matter fraction; WMF, white matter fraction; VVD, segmented ventricle volume difference; VBSI, ventricular boundary shift integral; ROI, region of interest.*

6.3.4 Discussion

In this study a direct comparison of methods for monitoring brain atrophy has been performed. Significant differences in atrophy rates and the variance of measurements were observed, suggesting that some methods may be more sensitive to global tissue loss

and would be a more effective marker of disease progression and treatment effects. In addition, some techniques appear to be more robust and provide more reliable and consistent results than others, an important factor if these methods are to be utilised in future clinical trials.

In contrast to previous investigations, this study found that measures of CCV were poor markers of brain atrophy, which appeared to be independent of the software used for analysis. CCV has previously been shown to decrease significantly over time in RRMS subjects, with changes ranging from -3.4ml $year^{-1}$ (approximately -1.1%) (Losseff *et al*., 1996) to -2.3cm^3 (approximately -0.8%), -1.4% (SD 2.3) and -2.6% over 18 months (Rovaris *et al*., 2000; Rovaris *et al*., 2001; Sailer *et al*., 2001). Atrophy rates in RRMS patients were much smaller in this study and there was an unexpected positive atrophy rate (brain "growth") in RRMS subjects when using Excalp. This was most likely due to the three outliers seen in this group, which also led to the large variance of measures seen with this technique. One limitation of the Excalp procedure that may have led to errors is that re-inclusion of brain tissue erroneously excluded by automatic processing is not possible. It was observed that even when baseline and repeat images were of similar quality, Excalp sometimes stripped more of one image than the other. Excalp also failed to derive a volume for one control subject suggesting that this software may not be consistently robust. In addition, it takes longer to derive measurements by Excalp than MIDAS as it is applied on a slice-by-slice basis.

Although MIDAS CCV gave a greater spread of values in controls and CIS than Excalp, both techniques detected highly positive atrophy rates in controls, again suggesting some degree of measurement error. It is likely that the greater variability may to some extent have been due to differences in slice thickness, slice selection and repositioning for follow-up images, leading to volumes of interest not necessarily being equivalent over serial imaging. Support for this idea comes from the measurements performed on registered images, which showed significantly less variability. However despite increased precision following image registration, the sensitivity of measurements and statistical power was still poor, and in controls brain "growth" of 0.4% $year^{-1}$ was detected. Although one study has previously found CCV changes of 0.2% over 18 months in control subjects (Rovaris *et al*., 2000), the finding in the current study is likely to have been the result of measurement error. Other studies of controls using different atrophy

measurement techniques have shown rates of around -0.2 to -0.3% year^{-1} (Fox *et al*., 2000b; Richert *et al*., 2006; Rovaris *et al*., 2005b; Scahill *et al*., 2003). One of the potential problems with registration of CSE images is the interpolation of voxels that is performed, which may not be entirely accurate, and this may have led to the errors observed in controls and the low atrophy rates detected in patients with CIS and RRMS.

Of all the methods tested, the BBSI and SIENA were shown to have the largest statistical power based on effect sizes calculated from atrophy rates in subjects with RRMS. However ventricular enlargement was also shown to have high statistical power and correlated well with whole brain atrophy rates measured by SIENA. In addition, analysis showed that the mean difference in ventricular enlargement when measurements were repeated on scan-rescan images was less than 0.1ml, and over one year VVD was the most consistent measure. There has been considerable variation in reported rates of ventricular enlargement, which may to some extent be due to analysis of different cohorts. In patients with RRMS with longer disease duration than the subjects studied in this investigation, median ventricular enlargement of 0.5ml year^{-1} (range -1.7 to 4.2) (Dalton *et al*., 2006), 2.1ml year^{-1} (IQR 0.7 to 3.7) (Fox *et al*., 2000b) and 2.3ml over six months (IQR -0.94 to 2.51) (Redmond *et al*., 2000) has been observed. Although CIS subjects were grouped together in this study, rates of ventricular enlargement were intermediate to those observed in a three year study of CIS patients who were divided into those who had developed MS at follow-up and those who had not; mean change over the three year study period was 2.4ml (95% CI 1.3 to 3.5) in patients who progressed to MS and 0.2ml (95% CI -1.0 to 1.4) in patients who did not progress (Dalton *et al*., 2004).

This is the first study to apply the VBSI to patients with MS and whilst there was no difference in the correlation of whole brain atrophy to the VBSI compared with VVD, the VBSI appeared to underestimate ventricular enlargement relative to VVD. This is a similar finding to that with whole brain BSI, and may be caused when the window over which the boundary shift is quantified does not necessarily include all the intensity changes that occur over the ventricular region (Fox & Freeborough, 1997). This discrepancy between the two measures also appeared to be greater when there was more ventricular enlargement. Periventricular lesions may have caused differences in the intensity transitions at the borders of the ventricles which can be accounted for when outlining regions in MIDAS, but may have caused some bias in VBSI measures. The

VBSI increased measurement precision relative to VVD however, and effect sizes were similar using the VVD and VBSI. One of the advantages of the VBSI when using only the baseline ROI is that analysis time would be reduced. This could be particularly useful in large multicentre studies where multiple assessments are performed. Following the segmentation of the baseline ventricular region, quantification of ventricular enlargement at each follow-up could be acquired automatically with minimal operator interaction. Although one might expect that measures using the VBSI would be more reliable and consistent over time, as the technique is automated, measurement reliability was slightly reduced relative to VVD and mean scan-rescan volume difference was greater. This may be because the method is more sensitive to small changes in scan acquisition. Although ventricular enlargement measures were shown to be sensitive, precise and reliable, it should be noted that the relationship between ventricular enlargement and brain atrophy may not always be consistent. In patients with Alzheimer's disease evidence was presented that the proportion of brain volume loss attributable to ventricular expansion is greater with increasing ventricular volumes. However this may not be true for different diseases, with different disease mechanisms and distributions of pathology (Schott *et al*., 2005).

SPM measures of WMF appeared to be subject to some degree of error, with inconsistent atrophy rates quantified between controls, CIS and RRMS subjects. Mean atrophy rates in control subjects were larger than might be expected, whilst a positive mean atrophy rate (i.e. increased WMF) was observed in patients with CIS. A similar finding has been observed previously; change in WMF was 0.2% over three years in CIS patients who had remained stable, and 1.3% in patients who had developed MS over the study period (Dalton *et al*., 2004). WMF atrophy rate in subjects with RRMS was similar to that found in a previous study that included 28 RRMS patients; -1.2% over three years was observed (Zivadinov *et al*., 2007). Evidence suggests that atrophy of the WM is inversely correlated with the volume of Gd-enhancing lesions (Tiberio *et al*., 2005), adding support to the hypothesis that inflammatory disease preferentially affects WM, which is likely to confound volume measures of this tissue compartment. This may partly explain the inconsistent results between controls, CIS and RRMS patients, and the relatively large variance of measures found in this study. However it may be that the distinction between WM and GM is inherently sensitive to noise in scan acquisition.

SPM measures of BPF and GMF gave more consistent results and the highest percentage changes of all the measures investigated, significantly larger in the case of GMF. The results from this study were in concordance with previous reports. Over three years a BPF change of -0.6% in CIS subjects, and -1.4% in subjects who had developed MS over the study period, was observed (Dalton *et al*., 2004). In subjects with RRMS a -2.5% change over three years has been reported (Zivadinov *et al*., 2007). In these same patients changes in GMF were -1.1%, -3.3% and -1.4% respectively. However the variance of these BPF and GMF measures in the present study was also high, and therefore statistical power was lower than that of the BBSI, SIENA and ventricular enlargement. Outliers in both the CIS and RRMS group, as observed in Figure 6-3, will have increased the variance but show that SPM may not be as robust as other techniques. SPM measures can be affected when the FOV is too small and there is either signal drop-off or the CSF is somewhat excluded. This may lead to misclassification of voxels and underestimation of CSF volume which may have been the cause of the outliers. It should also be mentioned at this point that SPM is an automated segmentation tool and therefore inaccurate tissue classification may occur, the degree of which can be difficult to assess. SPM also failed completely on one subject providing further evidence that it may not be as robust as other techniques. Although SPM may be useful to investigate tissue-specific changes that may help to elucidate disease mechanisms, lesion contouring must be performed which is time-consuming and a further disadvantage of this technique, especially for patients with high lesion loads.

Measurement reliability is important if atrophy rates are to be utilised in clinical trials where MRI acquisition may be performed at multiple centres. Previous studies have investigated scan-rescan reliability but these have been restricted to cross-sectional rather than longitudinal measurements. In addition the CV has usually been reported, making it difficult to compare results from this study with those reported previously (Carone *et al*., 2006; Fox *et al*., 2000b; Losseff *et al*., 1996; Paolillo *et al*., 2004; Rovaris *et al*., 2000; Zivadinov *et al*., 2004a; Zivadinov *et al*., 2007). Although SPM is an automated technique it appears that small differences in acquisitions may affect image segmentation, as evidenced by the ICC for repeated measurements of baseline to one year atrophy rates. The ICC was significantly lower for SPM measurements than all other measurements. Although the ICC was high for the BBSI, scan-rescan assessment estimated change significantly different from zero, which suggests some measurement error. This finding

may be due to the fact that the SD of measurements was the lowest of the techniques. Over the scan-rescan interval SIENA estimated the lowest change, suggesting it is robust to small differences in acquisition. In a cross-validation study of the BBSI and SIENA in which 185 scan-rescan image pairs were analysed, mean differences were 0.0006% for SIENA and 0.1118% for the BBSI, which are similar to those found in the current study (Smith *et al*., 2007). The ICC of repeated longitudinal measures was similar to that of the BBSI. To rigorously test measurement reliability and consistency of measures, it would have been valuable to have had scan-rescan images at the one year time-point also.

One of the limitations of this study is the multiple comparisons that were performed to investigate differences in the mean and variance of measures. Although these were limited to patients with RRMS, it is possible that significant effects were found by chance. However, in conclusion, there appear to be clear benefits to certain atrophy measurement techniques, namely the BBSI, SIENA and ventricular enlargement.

6.4 Chapter conclusions

This chapter has investigated the relative advantages and brain atrophy measurements of a number of techniques that have previously been applied to MRI of patients with a disease-related decrease in brain volume. It has been shown that registration of serial images increases measurement precision, which consequently increases statistical power and will allow better detection of disease progression and treatment effects. In particular the BBSI and SIENA, both registration-based methods, were shown to provide more precise brain atrophy measurements compared with techniques based on the subtraction of serial volumes. However ventricular enlargement was shown to correlate well with whole brain atrophy measures, and these measurements were highly reliable with relatively high statistical power.

Studies of larger cohorts with longer clinical and MRI follow-up are needed to investigate these measures further, and establish their potential as sensitive markers of the development of irreversible disability which may aid in identifying effective disease-modifying treatments and in prognosis for individual patients. Also, further work is needed to try and improve the stability of scan acquisitions.

7 Sample size calculations in relapsing remitting multiple sclerosis

7.1 Introduction

The standard primary outcome measures in trials of potential disease-modifying drugs in MS are clinically based and include the development of disability or relapse rate (Jacobs *et al*., 1996; Johnson *et al*., 1995). However, with evidence that neuroaxonal loss is a key pathological feature of MS (Peterson *et al*., 2001; van Waesberghe *et al*., 1999) which is widely considered to be the main pathological substrate of irreversible disability, it is likely that current scales which measure clinical disability may not reflect the extent or severity of this underlying and irreversible pathology. There may be a threshold of neuroaxonal loss only beyond which disability may become apparent, and pathology may occur in clinically silent locations which does not immediately cause a measurable change in clinical function. Moreover, it can be difficult to monitor the extent of irreversible disability in patients when using subjective clinical assessment scales that may not be sensitive enough to detect small changes in function, especially if the patient is experiencing relapses.

With the development of new disease-modifying treatments and potential neuroprotective agents for MS, there is an increasing need to monitor directly the efficacy of these drugs on the underlying global MS pathology, especially neuroaxonal loss. This thesis has investigated and optimised a number of atrophy measurement techniques that allow visualisation and measurement of the irreversible tissue loss that occurs as a result of this pathology and that may provide a means by which neuroaxonal loss can be inferred non-invasively in-vivo from structural MRI.

Currently, brain atrophy has only been used as a secondary (and retrospective) outcome measure in treatment trials (Filippi *et al*., 2004; Rudick *et al*., 1999), with the sample size having been determined for the primary outcome measure, e.g. relapse rate or disability for a phase III trial or MRI lesion activity for a phase II trial. If MRI measures of brain atrophy are to be adopted as markers of disease progression in future clinical trials of disease-modifying treatments for MS, the number of patients required to detect a given treatment effect should be determined. This is vital for any trial; if not enough subjects are

entered, there may be insufficient power to detect a treatment effect, whilst the inclusion of more subjects than necessary may expose additional subjects to harmful side effects. Both cases will ultimately result in wasted resources. This thesis has already identified several techniques that provide relatively precise, reproducible and sensitive measurements of brain atrophy, which may aid in reducing the number of patients required to detect a treatment effect.

In this study, brain atrophy was measured longitudinally at multiple time-points for up to three years using the three methods that were identified in Chapter 6 as being relatively more powerful markers of brain atrophy. Power calculations were then performed to determine the sample sizes required for a placebo-controlled trial with respect to the atrophy measurement method and duration of study. For sample size comparison, an atrophy measurement method that was deemed to be a less optimal marker of progression was also included.

7.2 Methods

7.2.1 Subjects and MR imaging

Sixteen control subjects and 33 patients with clinically definite RRMS (Poser *et al*., 1983) were identified from the cohorts included in this thesis (Chapter 3.1) (Table 7-1). Patients had experienced at least two clinical episodes, but were within four years of symptom onset (median 1.7, range 0.5-3.8 years). At study entry all patients had an EDSS ≤ 3 (median 1.5, range 0-3), and only patients with at least two MRI scans and who were not on disease-modifying treatment were included. The patients were representative of those who would be considered for inclusion in treatment trials. If patients were started on disease-modifying treatment, subsequent data was excluded from analysis.

T1-weighted coronal 3D FSPGR imaging was acquired at baseline and at up to six subsequent time-points (approximately six monthly up to 36 months) (Table 7-1), according to the protocols described in Chapter 3.3.

Table 7-1 *Characteristics of controls and subjects with RRMS who had available MRI data at each time-point for sample size calculations.*

		Time-point (months)						
		Baseline	***6***	***12***	***18***	***24***	***30***	***36***
Controls	***N (M:F)***	7:9	7:9	7:9	6:5	7:6	5:2	6:2
	Age (years) (mean, SD)	35.1 (6.3)	35.7 (6.4)	34.9 (8.4)	38.7 (6.6)	37.3 (7.1)	38.7 (9.0)	38.0 (5.4)
	Follow-up time (years) (mean, SD)	…	0.56 (0.06)	1.07 (0.08)	1.66 (0.24)	2.14 (0.18)	2.54 (0.07)	3.00 (0.09)
RRMS	***N (M:F)***	11:22	7:22	7:17	7:14	6:15	6:12	4:7
	Age (years) (mean, SD)	36.0 (7.4)	36.7 (7.6)	39.2 (7.4)	39.3 (6.5)	38.6 (6.5)	40.0 (6.4)	39.9 (6.8)
	Follow-up time (years) (mean, SD)	…	0.54 (0.06)	1.05 (0.07)	1.53 (0.10)	2.03 (0.12)	2.57 (0.16)	3.09 (0.20)
	EDSS (median, range)	1.5 (0-3)	2 (0-3.5)	1.5 (1-3.5)	1.5 (0-3.5)	2 (0-3.5)	2 (0-6)	2 (0-3)
	Disease duration (years) (mean, SD)	1.7 (0.8)	2.5 (0.8)	3.1 (0.8)	3.5 (0.8)	4.0 (0.7)	4.6 (0.7)	5.2 (0.8)

7.2.2 MRI analysis

(i) Segmented brain volume difference

Semi-automated segmentation of baseline and repeat brains was performed using MIDAS as described in Chapter 4.2.2, with correction of images for intensity inhomogeneity using N3 (Sled *et al*., 1998). Brain volume was determined from the segmentation within MIDAS.

(ii) BBSI

The BBSI was applied to quantify atrophy from each repeat image to baseline as described in Chapter 4.2.2. DBC (Lewis & Fox, 2004) was applied at registration and the BBSI window parameters were set at the optimal values determined in Chapter 4.3 (I_1=0.65, I_2=0.45).

(iii) SIENA

SIENA was applied to quantify atrophy from each repeat image to baseline as described in Chapter 5.3.2. This required axial reorientation of all images, and the initial brain segmentation was acquired using the optimised version of BET, described in Chapter 5.2.

(iv) Ventricular enlargement (VE)

Ventricle volumes were determined on all images as described in Chapter 6.3.2.

7.2.3 Statistical analysis

Separate linear mixed models for controls and patients were fitted to the data using Stata Xtmixed. The logarithm of the brain volume measurements against time from baseline was modelled. For BBSI and SIENA, the logarithm of the ratio of volume at the second scan to volume at the baseline scan was modelled, with the repeat volume calculated using the baseline volume and the BBSI/SIENA change (Frost *et al*., 2004). By modelling log volumes, these three models assume that brain volumes decrease proportionately with time. Conversely, for VE the absolute ventricular volume was modelled, consistent with an assumption of a constant volume increase with time.

Evidence of acceleration in atrophy rate was tested through the introduction of a fixed quadratic effect in time. In each model the repeated within-subject measurements were allowed for using random slopes, where estimation of the corresponding variance

component was possible. Random intercept effects were included in the models for BVD and VE. A further random effect was included in the BBSI and SIENA models to allow for the additional correlation structure inherent in direct measures of change (Frost *et al*., 2004).

Sample size calculations were performed for a trial including baseline and one follow-up assessment, using the estimated means and variances from these mixed models, and were based on the standard formula (Equation 7.1) with 90% power to detect a treatment effect and 5% two-tailed significance level. For BBSI and SIENA, calculations were based on performing an analysis of change, whereas for BVD and VE, calculations were based on an analysis of covariance. A completely effective treatment was considered to be one which reduced the atrophy rate to the level seen in controls, and an immediate and constant effect was assumed. The effect of different methods and trial duration on sample sizes was assessed by determining ratios of sample sizes, with 95% bootstrap CI calculated to indicate whether differences were statistically significant, treating the atrophy rate in controls as known.

$$Sample\ size\ per\ trial\ arm = \frac{(u+v)^2\ (2\sigma^2)}{(\mu_1 - \mu_2)^2} \qquad (7.1)$$

u = 1.28 to provide 90% power

v = 1.96 to test at the 5% significance level

μ_1 and μ_2 are the mean log brain volume atrophy rates (BVD, BBSI, SIENA) or mean VE rates in the placebo and treatment groups.

σ^2 is the variance of the ANCOVA log BVD atrophy rates, the variance of the log BBSI and SIENA atrophy rates or the variance of the ANCOVA VE rates.

The mean rate in the treatment group was taken as a percentage of the difference between control and RRMS rates. A completely effective treatment was considered to be one which reduced the atrophy rate to the level estimated in controls, i.e. a 30% reduction in atrophy rate in the treated group was equal to 30% of the difference in mean rate of the control and RRMS groups, which was subtracted from the mean RRMS atrophy rate.

A 5% drop-out of subjects and 5% of scan pairs being unusable was taken into account by dividing resulting sample sizes by 0.95^2.

7.3 Results

Eight RRMS patients started treatment with beta interferon during the study period; three patients by 12 months, and a further two by 18 months, one by 30 months and two by 36 months. Atrophy rates while on and off treatment were estimated, but there was no statistically significant evidence that these rates differed. It was therefore judged that bias to less disabled patients was not introduced by excluding on-treatment data.

Estimated atrophy in controls and patients with RRMS is shown in Table 7-2. For both controls and patients with RRMS, there was very little evidence of between subject variability in atrophy rate quantified by BVD (random slope effect could not be fitted). Consequently, the model predicted SDs of atrophy over one, two and three years were equal. There was no evidence for acceleration in mean atrophy rate over the three years in patients with RRMS using the BBSI ($p=0.90$), SIENA ($p=0.31$) or BVD ($p=0.08$). A linear rate of atrophy over this period was therefore assumed.

Estimated sample sizes for a parallel group, placebo-controlled design with atrophy rate as the outcome variable are shown in Table 7-3. For a given effect size, sample sizes were affected by measurement method, particularly at shorter trial durations. The smallest sample sizes were observed using SIENA, followed by VE, the BBSI and BVD. Over all trial durations sample sizes were statistically significantly smaller than BVD using the three other methods (Table 7-4). Although the BBSI, SIENA and VE showed similar sample sizes, SIENA gave marginally smaller estimates over all trial durations, which were statistically significantly smaller than the BBSI when considering trials conducted over one year. As expected, the longer the length of follow-up, the smaller the sample sizes required. For all atrophy measurement methods, statistically significant reductions in the sample sizes required were observed as the length of trial increased (Table 7-5).

Table 7-2 *Brain atrophy and ventricular enlargement over one, two and three year intervals, estimated from linear mixed models, in controls and subjects with RRMS. Results for the BBSI and SIENA take into account that the true atrophy rates differ from the mean atrophy rate.*

Interval	***Controls*** ***mean (SD) (95% CI)***				***RRMS*** ***mean (SD) (95% CI)***			
	BVD (% baseline volume)	***BBSI (% baseline volume)***	***SIENA (% baseline volume)***	***VE (ml)***	***BVD (% baseline volume)***	***BBSI (% baseline volume)***	***SIENA (% baseline volume)***	***VE (ml)***
1 year	-0.22 (2.27) (-0.59, 0.15)	-0.09 (0.39) (-0.21, 0.02)	-0.11 (0.30) (-0.22, -0.01)	0.11 (0.67) (-0.14, 0.35)	-0.84 (2.16) (-1.11, -0.57)	-0.63 (0.59) (-0.78, -0.48)	-0.79 (0.58) (-0.96, -0.61)	-1.51 (1.75) (0.98, 2.04)
2 years	-0.43 (2.27) (-1.17, 0.31)	-0.19 (0.46) (-0.42, 0.04)	-0.23 (0.42) (-0.45, -0.01)	0.21 (0.95) (-0.27, 0.69)	-1.67 (2.16) (-2.20, -1.13)	-1.26 (0.84) (-1.56, -0.95)	-1.57 (0.93) (-1.91, -1.22)	-3.02 (2.94) (1.96, 4.07)
3 years	-0.65 (2.27) (-1.75, 0.47)	-0.28 (0.57) (-0.63, 0.06)	-0.34 (0.56) (-0.67, -0.02,)	0.32 (1.30) (-0.41, 1.04)	-2.49 (2.16) (-3.28, -1.70)	-1.88 (1.13) (-2.33, -1.42)	-2.34 (1.33) (-2.85, -1.83)	-4.52 (4.24) (2.94, 6.11)

Table 7-3 *Estimates of the sample sizes required for parallel group placebo-controlled trials of a treatment reducing brain atrophy rate by varying degrees in patients with RRMS. Estimates are the numbers of patients required in each trial arm, with 90% power to detect a treatment effect at the 5% significance level, and taking into account a 5% subject drop-out rate and 5% of scans pairs being unusable.*

Trial duration	***Measurement method***	***Treatment effect size (% reduction in atrophy rate)***			
		30%	***50%***	***70%***	***90%***
1 year	*BVD*	3051	1098	560	339
	BBSI	314	113	58	35
	SIENA	191	69	35	21
	VE	269	97	49	30
2 years	*BVD*	763	275	140	85
	BBSI	157	56	29	17
	SIENA	123	44	23	14
	VE	140	50	26	16
3 years	*BVD*	339	122	62	38
	BBSI	128	46	23	14
	SIENA	111	40	20	12
	VE	114	41	21	13

BVD, segmented brain volume difference; VE, ventricular enlargement.

Table 7-4 *Relative sample sizes for each brain atrophy measurement technique. Relative sample sizes are not affected by power, significance level or effect size.*

	1 year			***2 years***			***3 years***		
	Relative sample size (95% CI)			***Relative sample size (95% CI)***			***Relative sample size (95% CI)***		
BVD	1	…	…	1	…	…	1	…	…
BBSI	0.10 * (0.06, 0.19)	1	…	0.21 * (0.09, 0.38)	1	…	0.38 * (0.16, 0.78)	1	…
SIENA	0.06 * (0.03, 0.13)	0.61 ** (0.37, 0.96)	1	0.16 * (0.07, 0.41)	0.79 (0.48, 1.28)	1	0.33 * (0.12, 0.86)	0.87 (0.49, 1.41)	1
VE	0.09 * (0.04, 0.21)	0.86 (0.56, 1.59)	1.41 (0.90, 2.88)	0.18 * (0.09, 0.43)	0.89 (0.55, 1.64)	1.13 (0.47, 2.44)	0.34 * (0.16, 0.82)	0.90 (0.42, 1.78)	1.03 (0.32, 2.47)

** p<0.05 for sample size ratio relative to BVD differing from 1; ** p<0.05 for sample size ratio relative to BBSI differing from 1; BVD, segmented brain volume difference; VE, ventricular enlargement.*

Table 7-5 *Relative sample sizes for different trial durations. Relative sample sizes are not affected by power, significance level or effect size.*

	BVD		***BBSI***		***SIENA***		***Ventricular enlargement***	
	Relative sample size (95% CI)		***Relative sample size (95% CI)***		***Relative sample size (95% CI)***		***Relative sample size (95% CI)***	
1 year	1	…	1	…	1	…	1	…
2 years	0.25 *	1	0.50 *	1	0.65 *	1	0.52 *	1
	(0.25, 0.28)		(0.36, 0.69)		(0.46, 0.86)		(0.45, 0.81)	
3 years	0.11 *	0.44 **	0.41 *	0.81 **	0.58 *	0.90 **	0.43 *	0.82 **
	(0.11 0.15)	(0.44, 0.52)	(0.24, 0.63)	(0.67, 0.92)	(0.36, 0.83)	(0.78, 0.97)	(0.34, 0.77)	(0.76, 0.95)

** p<0.05 for sample size ratio relative to one year; ** p<0.05 for sample size ratio relative to two years; BVD, segmented brain volume difference.*

7.4 Discussion

This study estimated the sample sizes required when using brain atrophy measurements from serial volumetric MRI to investigate treatment efficacy in MS. Results suggest that measurement of brain atrophy could be a practical addition to clinical endpoints for monitoring treatment effects on disease progression in phase III placebo-controlled trials of treatments for RRMS. However, it has been shown that it is vital for a sensitive and precise measurement technique to be used in order to minimise the number of patients required.

Studies have shown that it can take long follow-up periods to detect change using clinical outcomes (Paolillo *et al*., 2002; The IFNB Multiple Sclerosis Study Group, 1995). MRI-based measures of disease progression, notably lesion load quantification, have been investigated to determine their potential over clinical endpoints as a more objective and sensitive approach to assessing treatment efficacy in RRMS. One study looked at annual T2 lesion load increase and estimated that over a two year trial with 80% power to detect a treatment effect, 214 patients per treatment arm would be required to show a 30% reduction in the rate of increase (Molyneux *et al*., 2000b). This compares to an estimated 123 patients in each treatment arm in this study when using SIENA to measure brain atrophy. Of note is the difference in sample sizes estimated over shorter trial durations, with 638 patients required per treatment arm for a 30% reduction in lesion load over one year (Molyneux *et al*., 2000b), compared with 191 patients for a 30% slowing of atrophy rate (using SIENA).

Counts of new or enlarging lesions on monthly enhanced MRI may be more sensitive to treatment effects than T2 lesion load (Frank *et al*., 2004). Studies of monthly MRI over six months or less typically estimate sample sizes of less than 100 patients per treatment arm (Sormani *et al*., 2001; Tubridy *et al*., 1998). It appears however that to achieve statistical power greater than 80% with these numbers, treatment effects need to be in the order of 50% or more. Moreover, although this outcome measure may appear to require smaller patient numbers to show a treatment effect over shorter durations than brain atrophy, they are essentially assessing different aspects of MS. Whilst new enhancing lesions may reflect the acute stage of the disease, such lesions in general have shown limited correlations with future clinical disability (Fisher *et al*., 2002; Wolinsky *et al*., 2001). Brain atrophy is thought to have better pathological specificity for axonal loss and has been shown to correlate with later disease progression (Fisher *et al*., 2002; Paolillo *et*

al., 2002). It may be that counts of new and enlarging lesions are a suitable outcome measure for phase II trials, based on the sensitivity to detect treatment effects over short periods. However for longer phase III trials, and treatments hypothesised to prevent or slow neuroaxonal loss, brain atrophy is likely to be a more appropriate outcome measure.

In considering the feasibility of brain atrophy as an outcome measure, note that two year placebo-controlled phase III trials which have been performed in RRMS, with relapse rate or disability as the primary outcome measures, have involved a minimum of 100-200 patients per treatment arm (Jacobs *et al*., 1996; Johnson *et al*., 1995), and sometimes considerably more (Polman *et al*., 2006). These numbers equate to those estimated in this study for a 30% slowing in atrophy rate using SIENA, the BBSI and VE. In agreement with these findings, a study of interferon beta-1a in subjects presenting with a CIS suggestive of MS, found a statistically significant slowing of brain atrophy rate of approximately 30% over two years using SIENA, when comparing 123 subjects on treatment with 117 placebo patients (Filippi *et al*., 2004).

It was observed in Chapter 6 that measurement precision is greater using "direct" measures of atrophy, whereby serial brain images are positionally matched and the difference between them quantified, as opposed to subtracting absolute volumes. This study has shown the clinical consequences of this finding as the number of patients required to detect a treatment effect is influenced significantly by the method used to quantify brain atrophy. VE, BBSI and SIENA gave statistically significantly smaller sample sizes over all trial durations than BVD. Measurements of VE gave similar sample sizes to whole brain atrophy from SIENA and the BBSI. The benefits of this method have already been discussed, and these results provide further evidence that it should be considered as a suitable measure of brain atrophy and disease progression. However VE may not be specific to brain volume loss and could be influenced by factors such as hydration to a greater extent than whole brain atrophy measures. In addition it has been suggested that treatments could alter ventricular and brain volumes to different extents (Schott *et al*., 2005). However assessing alterations in the relationship of brain volume loss and ventricular expansion could provide information concerning treatment mechanisms.

All atrophy measurement methods gave statistically significantly smaller sample sizes the longer the trial duration, although the largest reductions were observed with segmented

volume difference as longer follow-up periods will usually compensate for lower measurement precision. It should be noted that a limitation of this study was the small number of subjects available at three year follow-up. This may have meant that estimation of the variance in atrophy rates at this time-point was poor, which in turn meant that there was less change in sample sizes estimated for a three year follow-up compared with a two year follow-up for the BBSI, SIENA and VE. In addition, whilst it may be advantageous to minimise the length of clinical trials to reduce costs, drop-outs and impact on patients, a phase III trial must be of substantial length to determine the efficacy and safety of the treatment. The hypothesised treatment effect must be considered fully, for example whether the treatment is effective immediately from administration and whether it has a constant effect over time.

Confounding factors that may influence brain atrophy measures such as demyelination, remyelination, gliosis, inflammation, oedema, dehydration and anti-inflammatory agents may need to be accounted for when designing trials. Studies have found that brain volume decrease was greater during the first year in patients receiving beta interferon than subsequent years, and it has been hypothesised that this is due to resolution of oedema initially (Hardmeier *et al*., 2005; Rudick *et al*., 2000). Although data acquired when patients were on treatment was excluded from the analysis, which could mean that the analysis is based on a patient group with milder disease symptoms, data was additionally analysed including patients on treatment, and no statistically significant differences in atrophy rates were found. In addition, the atrophy rates reported on patients who are not receiving disease-modifying treatment are comparable with those seen in other natural history and placebo RRMS cohorts (Rovaris *et al*., 2003; Rudick *et al*., 1999). Although the finding of no significant difference between on- and off-treatment groups could be taken as evidence that these patients need not be excluded from the analysis, it was felt that due to the small number of data-points that would be excluded it was better to maintain a more homogeneous group, given that treatment could be causing subtle unknown differences.

With disease-modifying treatments currently available for RRMS, there is an increasing possibility that new treatments must be tested against those currently available. Although this study has not addressed this aspect directly, larger sample sizes are likely to be required than for placebo controlled trials. However careful selection of patients recruited to clinical trials, focussing on those that have already shown progression in disability may

increase the power to detect treatment effects. In addition, data in this study was obtained and analysed in an identical fashion for all patients whilst for multicentre drug trials scanners, protocols and analysis may vary between sites and researchers. This could increase the variability in brain atrophy measurements relative to the current study and may increase the sample sizes required for a multicentre clinical trial. Note also that this study was based on three-dimensional volumetric MRI, whilst most studies and clinical trials in RRMS to date have investigated atrophy on scans with limited resolution in one plane. Whilst this study has only investigated the feasibility of using brain atrophy as a surrogate marker of disease progression in RRMS, it will be of particular relevance to determine the sample sizes required for trials of disease-modifying drugs in the progressive phase of MS when neuroaxonal loss may be more extensive.

Finally, although the sample sizes calculated in this study show that measurements of brain atrophy could be utilised to test treatment efficacy in RRMS patients participating in phase III clinical trials, recommendation that brain atrophy is the primary and definitive outcome measure is premature. This will only be possible when the relationship between atrophy and disability is better understood, and consistent evidence has emerged that the development of atrophy reflects and predicts disability.

8 Methodological considerations for longitudinal brain atrophy measurements: MRI scanner upgrades

8.1 Chapter introduction

Many of the techniques that have been developed to delineate regions and quantify atrophy rely on intensity differences between CSF, GM and WM. Over the duration of a longitudinal study, changes in the intensity profile (e.g. contrast between tissues) of an image could affect the reliability of segmentation and atrophy measurements. It is therefore important that the MR image is consistent in appearance and quality over time. Techniques that correct or normalise image intensity within and between serial scans (non-uniformity) have been developed that improve segmentation accuracy and precision (Lewis & Fox, 2004; Sled *et al*., 1998; Smith *et al*., 2002). However upgrades of scanner hardware and software can alter the quality and contrast of structural MR images in ways that cannot be corrected for using commonly available algorithms. Upgrades may be performed when changes have been made by the scanner manufacturer to improve image quality, decrease acquisition time, include additional features or increase the reliability of imaging. There are clearly many advantages to performing upgrades, but differences in the quality of images may be problematic when they occur during the course of longitudinal cohort studies. Whilst they may not necessarily affect qualitative visual assessment of MRI, they may influence quantitative atrophy measurements of the brain or smaller cerebral structures, especially when atrophy is subtle compared with the likely measurement error caused by scanner changes.

Whilst scanners invariably undergo upgrades, relatively few studies have reported the effect of these changes on brain atrophy measures. With brain atrophy being used increasingly in longitudinal trials of potential disease-modifying drugs and studies aiming to understand the progression of disease, it is important to determine the possible effect of routine scanner upgrades on these measurements. The aim of the first part of this study was therefore to determine the effect of a major scanner upgrade on volumetric measures of whole brain and lateral ventricles, derived using a semi-automated intensity threshold-based segmentation tool. In the second part of the study a potential method for correcting ventricular volume measurements for upgrade-related effects is explored.

8.2 Investigation into the effect of a major scanner upgrade on brain atrophy measurements

8.2.1 Methods

Subjects and MR imaging

Five healthy volunteers with no history of neurological complaints (three male, mean age at baseline 34.8 years (SD 7.8, range 24.9 to 43.3 years)) were scanned on a 1.5T GE Signa scanner, used for imaging of the other subjects included in this thesis, before and after a major scanner upgrade. Volunteers were members of staff at the Institute of Neurology and gave informed consent to be scanned. The upgrade included a change in the radio frequency transmitter coil, radio frequency amplifier, computer and software (Signa Horizon Echospeed 1.5T (5.8) pre-upgrade, and Signa EXCITE 1.5T (11.0) post-upgrade). Imaging was performed at regular intervals on nine or ten occasions during an average 5.3 month (SD 0.27) period prior to the upgrade, and on nine or ten occasions during an average 8.6 month (SD 2.3) period post-upgrade. Three of the five subjects were scanned using a coronal three-dimensional inversion-prepared FSPGR sequence with acquisition parameters TR=10.9ms, TE=4.2ms, TI=450ms, matrix 256x192, flip angle 20˚, FOV 240x180, resulting in 124 1.5mm thick slices. The other two subjects were scanned using an axial three-dimensional inversion-prepared FSPGR sequence with acquisition parameters TR=10.9ms, TE=4.2ms, TI=450ms, matrix 256x160, flip angle 20˚, FOV 300x230, resulting in 124 1.5mm thick slices.

MRI analysis

All scans were corrected for intensity inhomogeneity using N3 (Sled *et al*., 1998). Using MIDAS, measures of whole brain and lateral ventricle volume were obtained on all scans as described in Chapters 4.2.2 and 6.3.2. For ventricular segmentation, brains in MNI-152 standard space were registered to the earliest image acquired prior to segmentation.

Statistical analysis

Mean (SD) brain and ventricle volumes pre-and post-upgrade were calculated for each subject and a paired *t*-test was used to determine if there was a significant difference in volumes before and after the scanner upgrade.

Piecewise mixed effect multiple linear regression models (Equation 8.1) were fitted with outcomes whole brain volume and ventricular volume measured at different time-points on each subject, and as covariates an upgrade indicator (1=after upgrade, 0=before

upgrade), months centred on the upgrade date (i.e. taking value 0 at the time of the upgrade), and a month*upgrade interaction term. The coefficient for the latter term tests for before-after difference in rate of change, and is dropped in models where the upgrade does not affect this gradient. Subsequently the coefficient for 'months' estimates the mean upgrade-adjusted rate of change; models assuming no true change also drop this term. In all models the upgrade indicator estimates change due to the upgrade. These factors estimate 'fixed' effects, which characterise the average trajectories. The remaining terms of the model record the residual deviations from these averages, and are used to estimate the variabilities around the average:

$$volume_{ij} = \alpha_0 + \alpha_1.after_{ij} + (\delta_0 + v_j).u\_days_{ij} + \delta_1.u\_days*after_{ij} + u_j + e_{ij} \quad (8.1)$$

$volume_{ij}$=ith measurement of subject j
u_days_{ij}=time of ith measurement in subject j, measured in days centred on upgrade date
$after_{ij}$=upgrade indicator taking value 1 if the measurement is after upgrade, 0 if before
$u\_days*after_{ij}$=interaction term
α_0=estimated mean volume just before upgrade
α_1=estimated mean upgrade step, after – before
δ_0=estimated gradient of change in volume before upgrade: i.e. change in volume per day
δ_1=difference in gradient, after – before upgrade
u_j=random intercept, assessing between-subject variability in intercept
v_j=random slope, assessing between-subject variability in gradient
e_{ij}=within-subject variability around subject-specific means

8.2.2 Results

Figure 8-1 shows an example of the images obtained before and after the upgrade. Subtle differences in the intensity and contrast can be seen between the images. Table 8-1 gives the mean (SD) brain and ventricle volumes in each subject before and after the scanner upgrade whilst Figure 8-2 shows a plot of the pre- and post-upgrade brain and ventricular volumes in the five control subjects. There appears to be a consistent increase in brain volumes and a decrease in ventricular volumes following the upgrade. A paired *t*-test on these average pre- and post-upgrade volumes indicated that these differences were significant (both p=0.0012).

Figure 8-1 *T1-weighted volumetric images showing the differences in appearance when obtained a) pre-upgrade and b) post-upgrade. The images appear to be a slightly different intensity and contrast. The arrows indicate some areas where there is a noticable difference between the two images, most commonly darker more enlarged CSF spaces in the pre-upgrade image. The voxel intensities are directly comparable as the window and level settings within the viewing tool (MIDAS) are automatically adjusted between the two scans, by linearly scaling the contrast and width settings.*

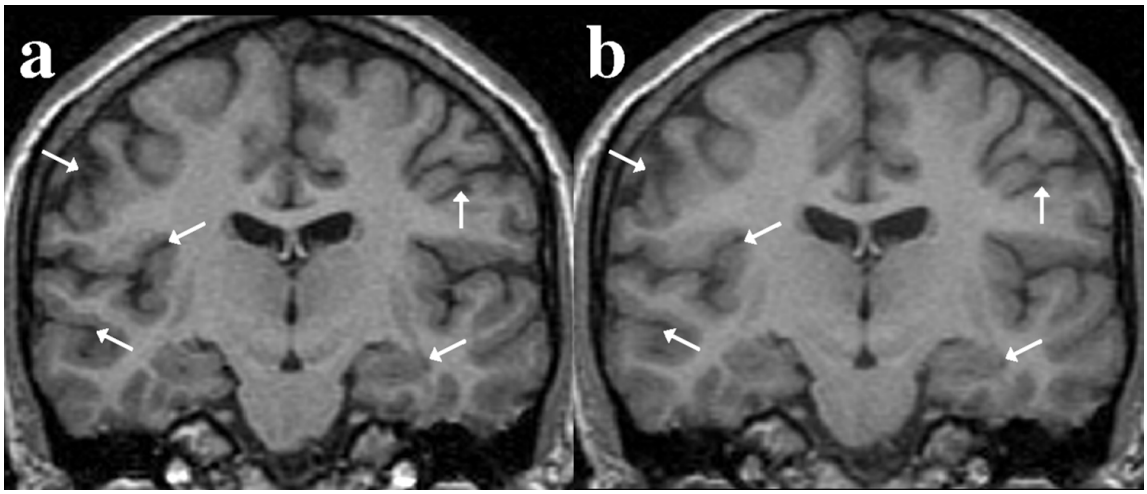
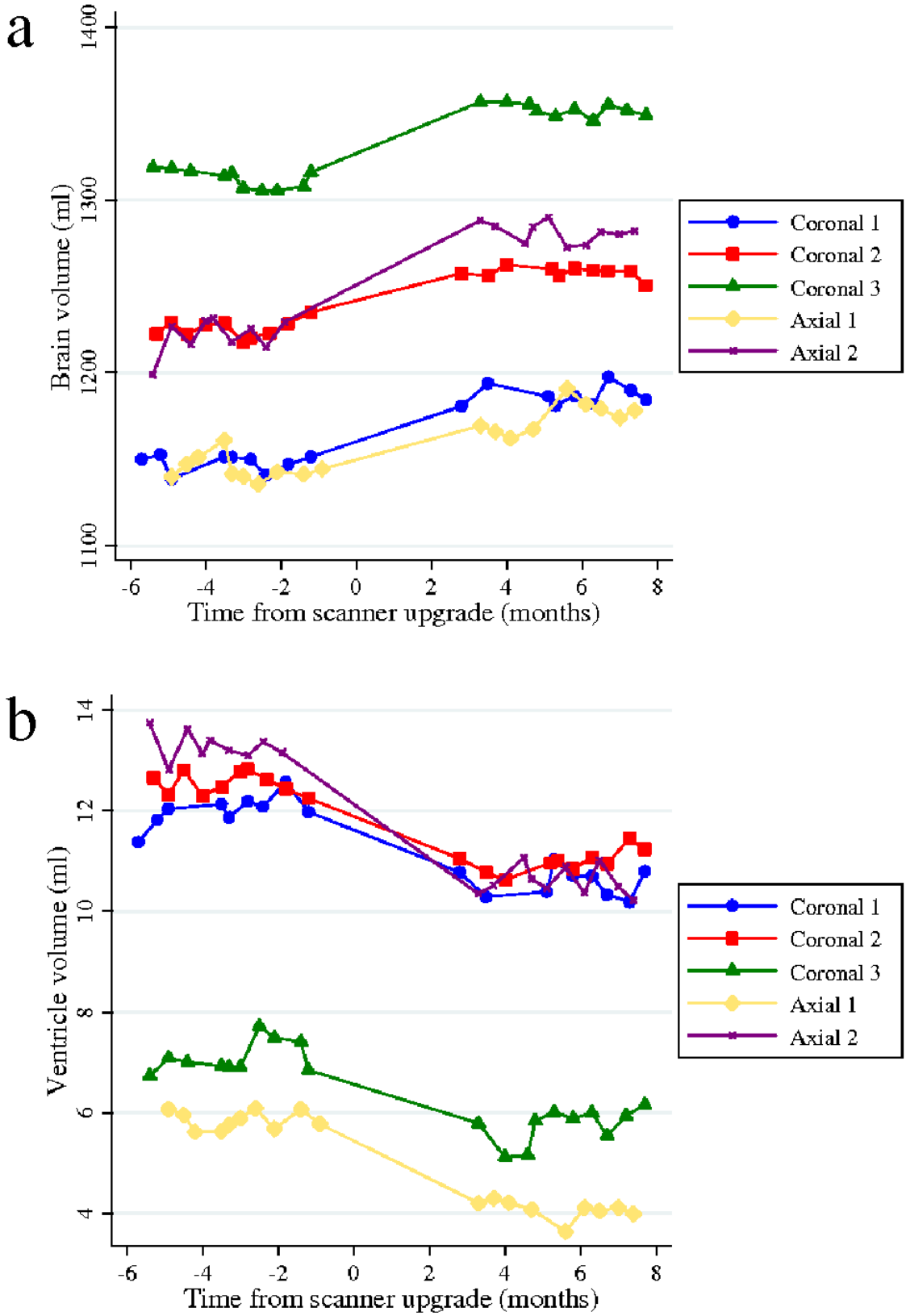


Table 8-1 *Mean (SD) brain and ventricle volumes pre- and post-upgrade in five control subjects scanned regularly before and after a scanner upgrade. Images were acquired in the coronal (three subjects) or axial (two subjects) plane.*

	Brain volume pre-upgrade (ml)	***Brain volume post-upgrade (ml)***	***Ventricle volume pre-upgrade (ml)***	***Ventricle volume post-upgrade (ml)***
Coronal 1	1148 (5.1)	1187 (5.9)	12.00 (0.32)	10.58 (0.29)
Coronal 2	1225 (5.3)	1258 (3.3)	12.55 (0.22)	11.00 (0.23)
Coronal 3	1312 (5.5)	1352 (3.7)	7.11 (0.32)	5.75 (0.36)
Axial 1	1144 (7.2)	1174 (9.1)	5.86 (0.18)	4.08 (0.19)
Axial 2	1221 (10.4)	1278 (10.7)	13.28 (0.28)	10.60 (0.29)

Figure 8-2 *Repeated a) brain and b) ventricle volume measurements in five control subjects scanned regularly before and after a scanner upgrade. Images were acquired in the coronal (three subjects) or axial (two subjects) plane.*



The graphs suggest that brain and ventricular volumes are stable over time except for the "jump" in values over the upgrade. This is to be expected as the upgrade will not change the effect of aging and is also unlikely to alter the scanner stability. However as the sample was small, this was tested for using the model. For brain volumes, the model estimated the difference in gradient of measures pre- and post-upgrade to be -0.0021 (95% CI -0.064 to 0.068, p=0.95). For ventricular volumes, the estimated difference in gradient pre- and post-upgrade was also small and non-significant: -0.00012 (95% CI -0.00272 to 0.00248, p=0.93). Therefore it was assumed that there was no change in the gradient for either brain or ventricular volumes and this term was dropped from the model. Subsequent reported results assume a common gradient before and after the upgrade, with only a step change due to the upgrade.

Assuming a gradient in measured volumes, the model estimated the upgrade-related brain volume change to be 38.66ml (95% CI 29.6 to 47.7, p<0.001). The estimated gradient was 0.005ml day^{-1} (95% CI -0.042 to 0.053, p=0.83). As there was no evidence to suggest a statistically significant gradient of change, it is sensible to re-run the model under the assumption that the small gradient is not real, and that there is in truth no change. Under this assumption the volume change related to the upgrade is slightly increased; a model excluding the gradient term estimated the upgrade-related brain volume change to be 40.15ml (95% CI 36.8 to 43.5, p<0.001). Of note is the fact that this model assuming no gradient in measured volumes is essentially equivalent to, though more efficient than, the paired *t*-test, and therefore confirms the finding of a statistically significant upgrade-related change in volume. Based on the model-estimated mean brain volume of 1210.56ml prior to the upgrade (α_0), the upgrade-related volume change as a percentage of total brain volume was 3.3%.

Similarly for ventricular volume, assuming a gradient in measured volumes, the model estimated the upgrade-related ventricular volume change to be -2.01ml (95% CI -2.36 to -1.65, p<0.001). Again, the estimated gradient of volumes was small and there was no evidence to suggest that it was statistically significant; 0.00095ml day^{-1} (95% CI -0.00119 to 0.00310, p=0.39). Excluding the gradient term from the model led to a decrease in the estimated upgrade-related change in ventricular volume; -1.76ml (95% CI -1.90 to -1.61, p<0.001). This effect was the opposite to that seen in the brain volume model because the overall gradient is in the opposite direction to the upgrade step. Based

on the model-estimated mean ventricle volume of 10.15ml prior to the upgrade (α_0), the upgrade-related volume change as a percentage of total ventricle volume was -17.3%.

Assuming a small gradient in volumes, estimates of the variability within and between subjects are shown in Table 8-2. Relatively small within-subject variability was detected and there was no evidence that this variability differed pre- and post-upgrade (brain volumes p=0.74, ventricular volumes p=0.66). The variance in mean volumes was the greatest contributor to between subject variability.

Table 8-2 *Linear regression model-derived estimates of the variability in brain and ventricular volumes between and within control subjects, scanned regularly before and after a scanner upgrade, when assuming a small gradient in volumes over time.*

		Brain volumes (variance, 95% CI)	***Ventricle volumes (variance, 95% CI)***
Estimated between-subject variation	***Slope***	0.0015 (0.0003,0.0069)	0.0000038 (0.0000009, 0.000016)
	Intercept	4943.4 (1235.5, 19779.4)	11.16 (2.79, 44.66)
	Covariance [correlation] between slope and intercept	0.71 [0.26] (-2.19, 3.61)	-0.0023 [-0.34] (-0.0092, 0.0047)
Estimated within subject variation		44.4 (32.9, 60.0)	0.069 (0.051, 0.093)

8.2.3 Discussion

This study has shown that MRI scanner upgrades can have a significant effect on MRI-derived volumes of the brain which are dependent on intensity threshold-based delineation of this structure from CSF. Such effects are significant enough to influence brain atrophy measures over the course of longitudinal studies crossing an upgrade. These results suggest that ideally scanner upgrades should be avoided during the course of longitudinal MR studies of quantitative brain volume measurements. However this may be impractical given that scanners are used for multiple studies with overlapping time frames. The statistical model applied in this study may be useful to compensate for changes in acquisition without the need for time-consuming post-processing of images. This should reduce the amount of unusable data in future longitudinal studies.

In this study an upgrade-related increase in brain volume and a decrease in ventricular volume were observed. Comparing images obtained pre- and post-upgrade a noticeable change in image contrast and an increase in intensity had occurred. This affected measurements of brain and ventricular volume which rely on the contrast between brain and CSF. The border regions of the brain in the post-upgrade image included more voxels of a higher intensity than the pre-upgrade image, and therefore more voxels were defined as "brain" and fewer as "CSF". This is in spite of "normalising" brain intensity over the mean intensity of the whole brain, and thresholds for ventricular segmentation being set as a percentage of brain intensity.

Although the change in contrast and intensity between pre- and post-upgrade images was subtle when scans were inspected visually, the associated change (increase) in brain volume was estimated to be approximately 3.3% of WBV. This increase is typically much greater than the annual atrophy rates (e.g. decreases of 0.5-2%) seen in neurological disorders such as MS (Rovaris *et al*., 2003), Alzheimer's disease (Schott *et al*., 2005), and Huntington disease (Henley *et al*., 2006). Similarly, the upgrade-related decrease in ventricular volume was approximately 2.0ml (17.3% of estimated ventricular volume), which is within the range of annual increases that have previously been observed in these diseases (Dalton *et al*., 2006; Schott *et al*., 2005). It is also of note that the upgrade-related changes observed in this study are in the opposite direction to biological atrophy changes.

In contrast to the current study, a recent investigation found no noticeable bias in cortical thickness measurements from images obtained over a major scanner upgrade (Han *et al*., 2006). In fact the upgrade improved the reliability of measurements post-upgrade, probably reflecting the increased SNR. Another study, investigating brain atrophy in patients with Alzheimer's disease and controls, found that a major hardware change between scan pairs had a negligible effect on the observed atrophy rates quantified using two registration-based methods, and did not decrease the group separation between controls and patients (Gunter *et al*., 2003). The difference in findings between these studies and the current one may be due to several factors. Firstly, the actual procedures that were performed at the upgrades are likely to have varied, which could cause different changes to acquisitions. Secondly, different acquisition parameters and pulse sequences were used in the three studies, which may have been affected differently by the associated upgrade. Lastly, these studies all used different atrophy measurement methods that may be relatively more or less robust to changes in image quality. Methods that rely on

intensity and contrast differences between brain and CSF are more likely to be affected. It would be useful to investigate the effect of the upgrade on other measures of brain atrophy used in this thesis, such as those which measure brain volume loss directly following the registration of serial images, e.g. SIENA (Smith *et al*., 2002) and the brain and ventricular BSI (Freeborough & Fox, 1997).

With brain atrophy measurements being used increasingly to monitor disease progression in a number of neurological conditions, it is important that potential upgrade-related effects on these measures are identified and corrected for if found to be significant. This may be important for relatively short-term studies taking place over one to three years when it is desirable to reduce measurement errors that may confound outcomes; disease-related brain atrophy changes over such periods are small, and in trials of potential disease-modifying treatments real treatment effects could be obscured by larger magnitude spurious changes due to an upgrade. It is also important in longer-term follow-up studies of subjects, which may wish to investigate the progression of atrophy and its relationship to clinical outcomes and other MRI markers. For example a study that investigated atrophy over eight years in MS patients, initially recruited for a two year treatment trial, analysed BPF (Fisher *et al*., 2000). Although MRI at eight years were acquired with pulse sequence parameters as similar as possible to the original study and progressive brain atrophy was detected, there is likely to have been subtle changes in the images that may have affected outcome measures.

Several strategies could be utilised to minimise upgrade-related changes over longitudinal studies. Firstly, changes in the acquisition parameters and pulse sequences can be made to obtain an image as close as possible in appearance to pre-upgrade images. However this will reduce the advantages of performing the upgrade in terms of obtaining images with greater CNR and SNR, and other benefits. Secondly, post-processing of the images could be performed that could more closely match images in appearance pre- and post-upgrade. Thirdly, statistical modelling of data obtained from pre- and post-upgrade images can be performed and used to correct for artificial volume changes. This method will allow more realistic changes in MRI measures to be obtained over upgrades for use in future studies, whilst minimising time-consuming post-processing methods. Techniques that do not require absolute or relative intensity values, such as some registration-based methods, to determine volume change may be less affected by alterations in image contrast changes and should be investigated in this respect.

The statistical model included volume measurements from control subjects at multiple time-points, and therefore allowed accurate and reliable estimates of the upgrade step and gradients of change pre- and post-upgrade. However this model could be adapted to be used in datasets of subjects in whom brain atrophy may be occurring more rapidly (involving the gradient terms of the model to a greater extent), and when only single scans are available for subjects pre- or post-upgrade. Adaptations to this model for these purposes will need to be investigated further. Although the gradient of measures was small and non-significant in this control cohort, it could be misleading not to include this term in models of patient data, particularly over longer studies in neurodegenerative diseases, where small gradients are expected due to real volume changes over time. Excluding the gradient factor from the model in the current study altered the upgrade-related volume change estimate marginally.

In summary, it has been shown that changes in image acquisition related to scanner upgrades can occur, despite all attempts to maintain identical acquisition parameters, and the effects of these changes need to be taken into account. Other factors may also impact on MRI-derived volume and atrophy measures in longitudinal studies, e.g. changes in acquisition parameters, subjects being scanned on different scanners, scanner drift, subject positioning and subject hydration. Regular scanning of control subjects prior to and following planned upgrades will help to identify upgrade-related changes which it may then be possible to adjust for when investigating true longitudinal biological changes.

8.3 Can a new protocol for ventricular segmentation correct for upgrade-related changes in ventricle volume?

8.3.1 Introduction

The analysis of control images over an upgrade in the previous study demonstrated that significant changes in volume measurements can occur due to scanner upgrades. These must be corrected for if reliable quantitative measurements of brain volume and atrophy rates are to be obtained. A statistical model for simultaneous estimation and correction of upgrade-related changes was described in the previous section, but it may be that for cross-sectional measurements of volume, simple alterations to segmentation methods could be performed which might effectively correct measures. This would allow correction of upgrade-related effects on volume when specialist statistical help is unavailable or when investigating single subjects, when it is not possible to generate a regression model. Given the observed changes on visual inspection and through further

examination of the images, it was determined that although mean brain intensity was similar on pre-and post-upgrade images, CSF was darker on the pre-upgrade scans relative to post-upgrade scans. Given the fact that many segmentation methods, including the brain and ventricular segmentation methods used in the previous section, are based on thresholding and rely on image contrast, it was hypothesised that by altering the protocol (the percentage of mean brain intensity used as the upper threshold) for ventricular segmentation on post-upgrade images, "upgrade-corrected" volumes could be achieved. The aim of this study was therefore to determine the optimum percentage of mean brain intensity for ventricular segmentation on post-upgrade images, and to test whether ventricular segmentation based on this threshold method did indeed correct for upgrade-related changes.

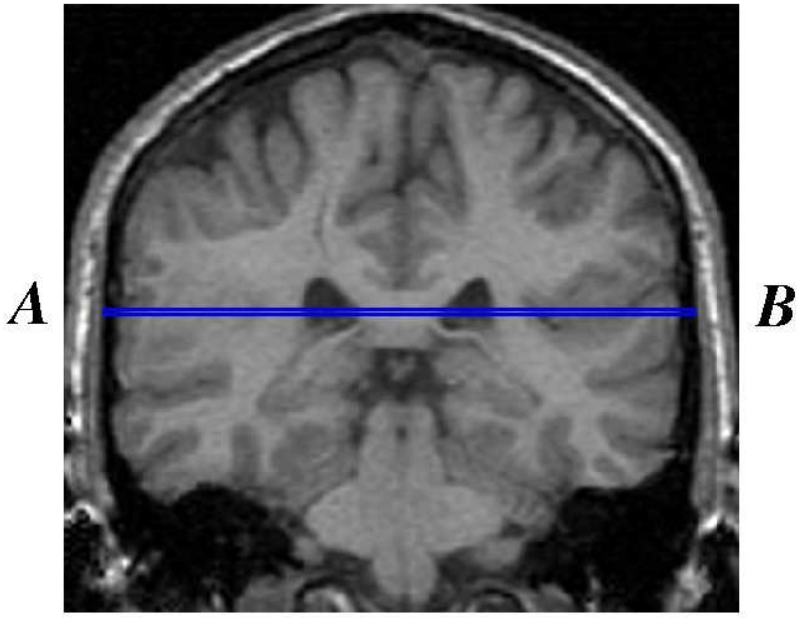
8.3.2 Methods

Theory behind alteration of the segmentation protocol

Due to the difference in CSF intensity between the pre- and post-upgrade images, there is less intensity change across the brain/CSF border on post-upgrade images compared with pre-upgrade images. This is demonstrated in Figure 8-3 which shows the intensity profiles of representative pre- and post-upgrade images from one subject. Using the standard upper threshold value of 60% of mean brain intensity for ventricular segmentation, voxels which have an intensity above this value are excluded. As the post-upgrade scan has more voxels of a higher intensity in the brain/CSF boundary region, this means a greater number of voxels are excluded from the ventricular segmentation compared with the pre-upgrade image, making the ventricular region smaller on the post-upgrade scans (assuming there is no real change in volume) (Figure 8-4a). Using a higher threshold value (as a percentage of mean brain intensity) could compensate for the generally higher intensity of the voxels in the border region, by including voxels with intensity >60% (and <x%) of the mean brain intensity (Figure 8-4b).

The value that the new threshold for ventricular segmentation on post-upgrade images should take can be calculated as a percentage of the gap between CSF and brain intensity and so the aim is to match these between pre- and post-upgrade images (Equation 8.2). The new value can be expressed as a percentage of mean brain intensity for application to other images (Equation 8.3).

Figure 8-3 *Intensity profiles of a pre-upgrade and a post-upgrade T1-weighted image from one subject. A one-dimensional intensity profile through the brain on each image was generated by drawing a line one voxel high between the points A and B (top image), and normalising the intensity value of each voxel in this region to the mean whole brain intensity. The area where the region of interest crosses the ventricles is labelled on the intensity profile.*



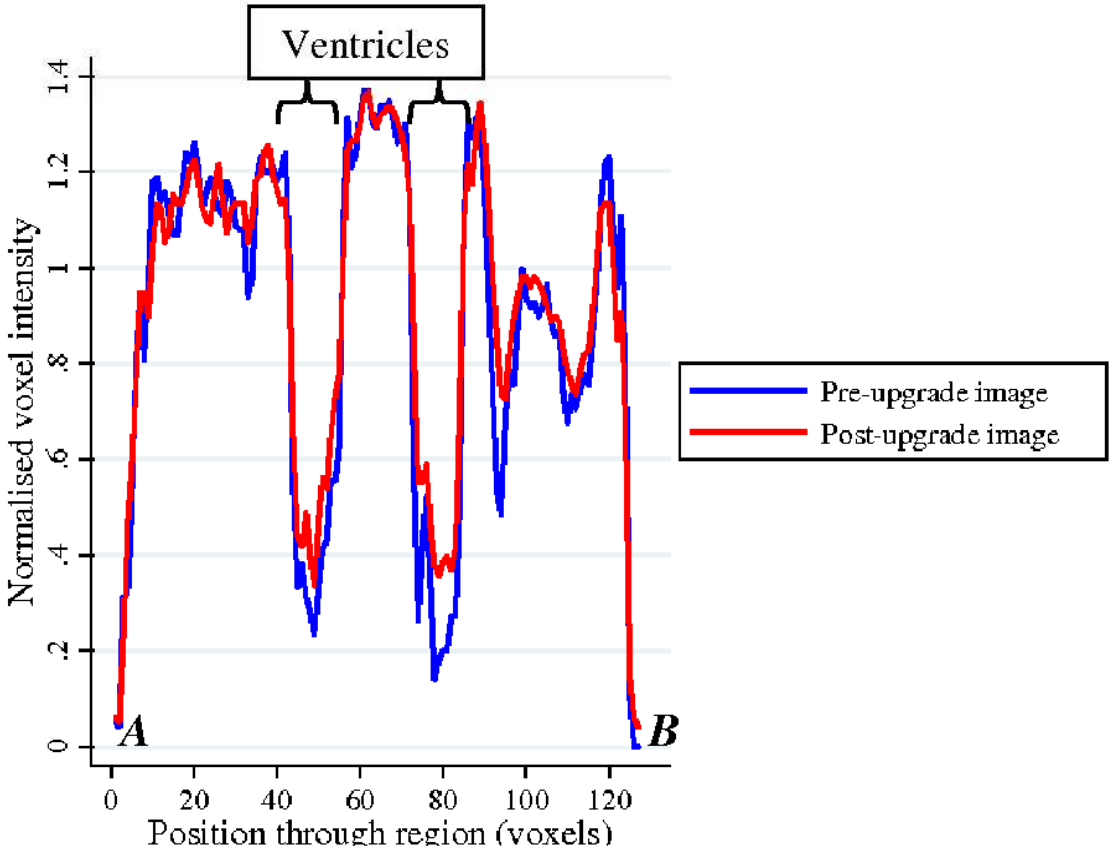
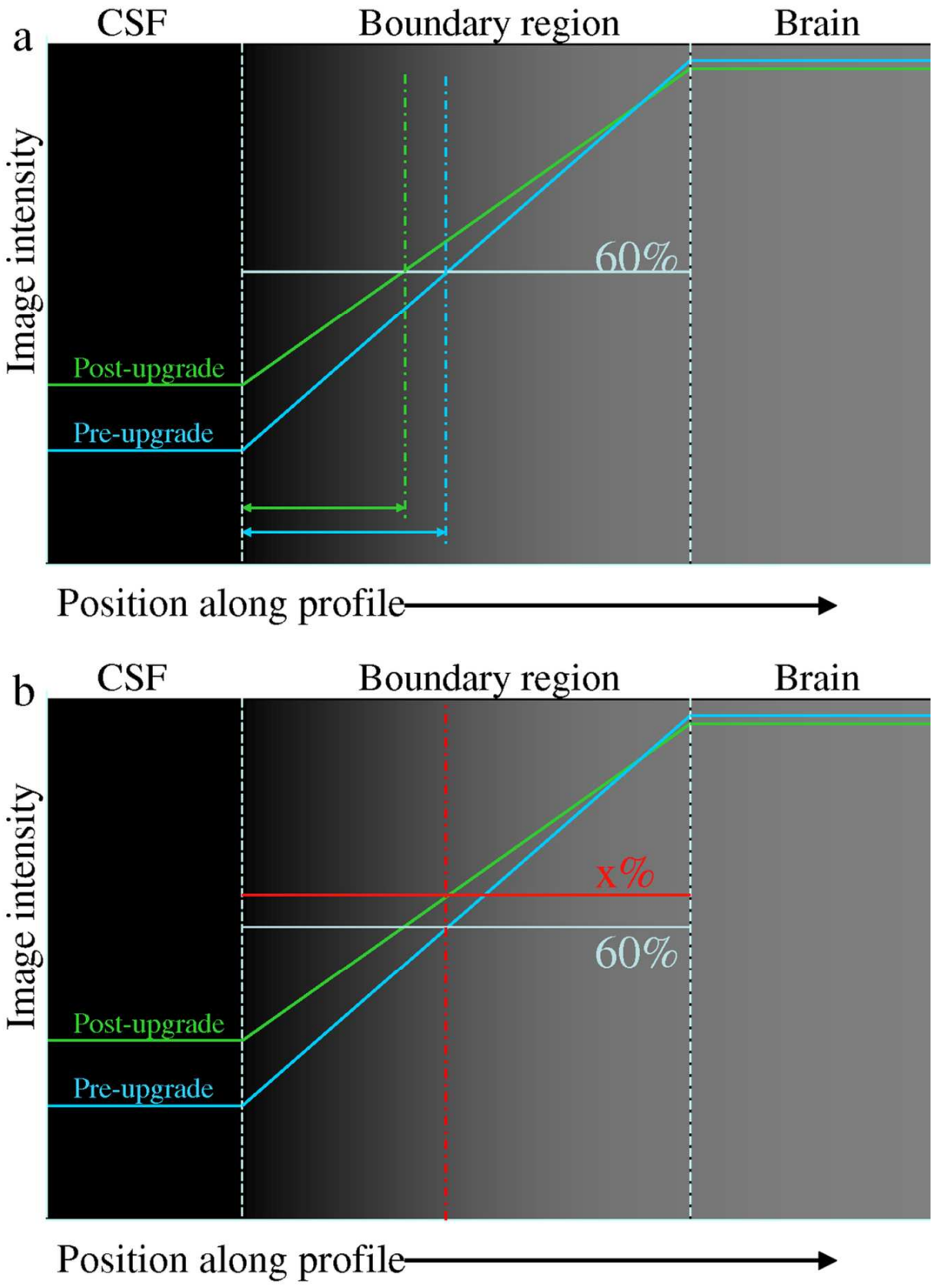


Figure 8-4 *Effect of differences in image contrast due to a scanner upgrade on ventricular segmentation and quantification of volume. The figures show one-dimensional representations of the intensity profiles through a brain boundary for pre- and post-upgrade images, and assume no real change in ventricle volume. a) The 60% upper threshold used for ventricular segmentation and the corresponding ventricular boundaries on pre- and post-upgrade images (-.-.-.-), b) Using a higher threshold (x%) for the post-upgrade image, the same boundary region can be achieved on the post-upgrade image as when using a 60% threshold for the pre-upgrade image*



$$\frac{T_{pre} - I_{csf(pre)}}{I_{brain(pre)} - I_{csf(pre)}} = \frac{T_{post} - I_{csf(post)}}{I_{brain(post)} - I_{csf(post)}} \quad (8.2)$$

$$x\% = \frac{T_{post}}{I_{brain(post)}} \text{x}100 \quad (8.3)$$

T=threshold, I=mean intensity, pre=pre-upgrade image, post=post-upgrade image.

Subjects and MR imaging

The five control subjects and registered images included in the previous experiment were utilised in this study (see Chapter 8.2.1).

MRI analysis

The first post-upgrade image available for one of the subjects was selected and the lateral ventricles were segmented using an upper threshold of 68% of mean brain intensity. Based on this ventricular region, and the brain region obtained for this image in Chapter 8.2.1, new estimates of mean brain and CSF intensity were determined and the values entered into Equation 8.2, in conjunction with intensity values for the pre-upgrade image of the same subject (the last image before the upgrade). Based on the result, it was determined whether the percentage threshold needed to be increased or decreased and the segmentation performed again, changing the threshold by one point. This process was repeated iteratively until the optimal threshold that most closely matched pre- and post-upgrade images according to Equation 8.2 was reached. The percentage of mean brain intensity that this value represented was then determined according to Equation 8.3, and the ventricles segmented on all other post-upgrade images for the five subjects, based on this percentage threshold.

8.3.3 Results

The optimum percentage threshold for ventricular segmentation on post-upgrade images was calculated to be 67%. Figure 8-5 shows a plot of the ventricular volumes in the five control subjects using this new threshold for post-upgrade images, whilst Table 8-3 gives the mean (SD) ventricle volumes in each subject before and after the scanner upgrade. Average volumes appear to be more constant over the upgrade using the new threshold, and there was no evidence that mean values pre- and post-upgrade were significantly different (p=0.98).

Figure 8-5 *Repeated ventricle volume measurements in five control subjects scanned regularly before and after a scanner upgrade, using 67% of mean brain intensity as the upper threshold for ventricular segmentation on post-upgrade images. Images were acquired in the coronal (three subjects) or axial (two subjects) plane.*

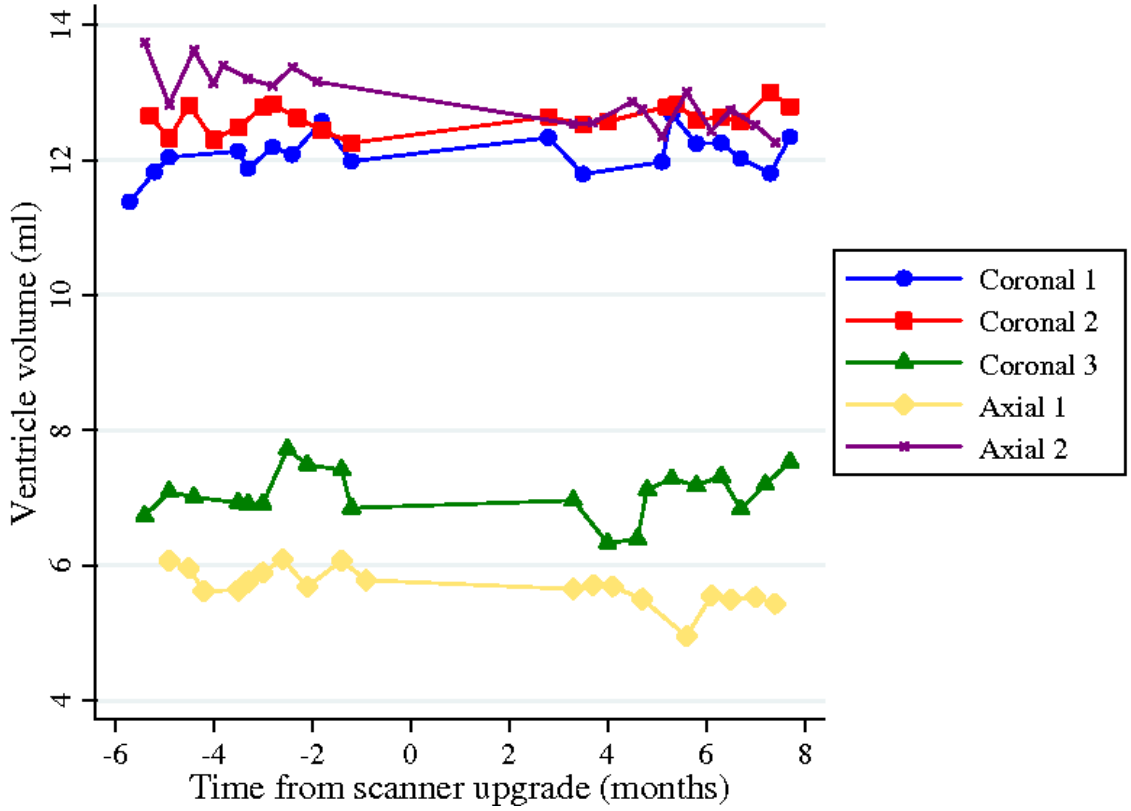


Table 8-3 *Mean (SD) ventricle volumes pre- and post-upgrade in five control subjects scanned regularly before and after a scanner upgrade, using 67% of mean brain intensity as the upper threshold for ventricular segmentation on post-upgrade images. Images were acquired in the coronal (three subjects) or axial (two subjects) plane.*

	Ventricle volume pre-upgrade (ml)	***Ventricle volume post-upgrade (ml)***
Coronal 1	12.00 (0.32)	12.16 (0.29)
Coronal 2	12.55 (0.22)	12.69 (0.15)
Coronal 3	7.11 (0.32)	7.01 (0.39)
Axial 1	5.86 (0.18)	5.50 (0.23)
Axial 2	13.28 (0.28)	12.60 (0.23)

As the sample was small, these results were tested using the statistical model described in the previous investigation (see Chapter 8.2.1). The model estimated the difference in gradient of measures pre- and post-upgrade to be -0.00046 (95% CI -0.0030 to 0.0021,

p=0.72). Therefore it was assumed that there was no change in the gradient and this term was dropped from the model. Assuming a gradient in measured volumes, the model estimated a change in ventricle volume over the upgrade of -0.40ml (95% CI -0.76 to -0.05, p=0.025), despite using the altered threshold for ventricular segmentation on post-upgrade images. The estimated gradient was 0.0018ml day^{-1} (95% CI 0.00048 to 0.0032, p=0.008). The volume change related to the upgrade as a percentage of total ventricle volume was -3.7% (based on a mean ventricle volume of 10.78ml).

8.3.4 Discussion

This study investigated a simple method for correction of upgrade-related changes in measured ventricle volumes, which required estimation and application of a new intensity threshold with which to segment the ventricles on post-upgrade images. Although there was no significant difference in mean ventricle volume before and after the upgrade when tested for using a paired *t*-test, the more stringent multiple linear regression model which accounted for a gradient in measured volumes found that ventricle volumes post-upgrade were significantly smaller than pre-upgrade values.

There are several possible reasons why this method may have failed to adequately correct for the upgrade-related changes. Firstly, this method assumes that the intensity profile over the brain/CSF boundary is linear. This may not be true if there are local subvoxel fluctuations in intensity due to artefacts for example. Secondly the threshold value can only be set as an integer, which may account for some variability between pre- and post-upgrade volumes. The range of intensities over the boundary was relatively small, but the intensities could be scaled over the entire image to allow a greater range of values however. In addition, it is not clear whether ventricular size could affect the method, for example larger ventricles may have more partial volume voxels that are included using the new method compared with smaller ventricles, therefore causing artificially greater enlargement in subjects with bigger ventricles.

As such these results suggest that the proposed method does not completely correct for upgrade-related changes in ventricle volume measured using MIDAS. It is possible that a similar approach to that presented in this study could be useful for other segmentation and volume quantification techniques based on methodology similar to that used for ventricular segmentation. This would have to be investigated further, but could allow correction of volumes from a single subject and time-point.

8.4 Chapter conclusions

In summary, this chapter has shown the significant impact major scanner upgrades can have on brain volume measurements. Due to a scanner upgrade, longitudinal measurements of brain and ventricular volume showed changes in the opposite direction to those expected when brain loss and normal aging occur. Whilst not investigated in this study, it is likely that regional measures (e.g. the cortex, caudate, CC) will be similarly affected by scanner upgrades. Only volumetric FSPGR images acquired over one particular scanner upgrade were investigated in this study, however it is likely that most major scanner upgrades, will have some effect on brain volume measurements. Even minor scanner upgrades could have subtle effects, and regular scanning of phantoms and control subjects should therefore be routinely performed to monitor these effects.

These upgrade-related changes need to be corrected for if reliable longitudinal analysis is to be performed. Although a simple method for correction of post-upgrade ventricular volumes was investigated, it did not adequately compensate for the upgrade-related changes. For large longitudinal studies, statistical methods may prove to be the most reliable approach with which to correct brain volumes and atrophy measurements. In this chapter a regression model was presented that can be used to estimate upgrade-related changes in volume. Simultaneous estimation and correction of post-upgrade volumes in control and patient cohorts using this model must now be investigated. In addition, investigation into the ability of this model to correct volumes when only one image is available pre- or post-upgrade must be performed.

9 Relationship of brain atrophy to clinical progression in patients with clinically isolated syndromes and relapsing remitting multiple sclerosis

9.1 Chapter introduction

It has been identified that certain techniques may provide better measures of brain atrophy than others, and that the sensitivity and precision of measurement can be optimised to allow brain atrophy, and potentially disease progression, to be tracked more efficiently. Although the previous chapters have confirmed that brain atrophy occurs from the very earliest stages of the disease, it is still necessary to determine how brain atrophy relates to the progression of clinical disability in these patients. In addition, cognitive function is often affected in patients with MS and the association of brain atrophy to this aspect of the disease must also be investigated. Long term follow-up studies are required to explore the relationship of progression of brain atrophy to clinical disability and cognition. This chapter investigates such relationships through five year clinical follow-up of patients presenting with CIS and with RRMS.

9.2 Investigation into the prognostic value of brain atrophy rate for clinical progression in subjects presenting with a clinically isolated syndrome suggestive of multiple sclerosis

9.2.1 Introduction

In Chapter 6.2 it was shown that brain atrophy rate measured over the first year following initial symptoms was significantly greater in CIS subjects who had progressed to MS by three year follow-up, than in controls and subjects who remained CIS. This confirmed previous reports that have shown ventricular enlargement, and changes in BPF and GMF, to be greater in CIS patients progressing to MS over periods of up to three years (Brex *et al*., 2000; Dalton *et al*., 2004).

Currently, it is known that the probability of progression to a diagnosis of MS in CIS subjects is greater in those subjects who have visible T2-weighted lesions on MRI acquired at presentation (Barkhof *et al*., 1997; Brex *et al*., 2002; Dalton *et al*., 2002a). However with evidence that brain atrophy is an early feature of MS, it is possible that this could provide an additional indicator of prognosis in subjects presenting with CIS. With

research showing a beneficial role of early immunomodulatory treatment in patients presenting with CIS (Comi *et al*., 2001; Filippi *et al*., 2004; Jacobs *et al*., 2000; Kappos *et al*., 2007), the identification of patients with a high risk of developing MS, and those who may have a faster evolving disease course and would be likely to benefit most from treatment is of clinical relevance. The aim of this study therefore was to investigate whether early brain atrophy rate is independently related to the risk of development of MS. One of the limitations of the investigation of CIS in Chapter 6.2 was that subjects had been followed-up for only a three year period, and that they may have been diagnosed with MS at a later date. Therefore the follow-up period of CIS patients in this study was extended, and only subjects that had a clinical assessment at five year follow-up were included.

9.2.2 Methods

Subjects and MR imaging

Thirty-seven subjects (14 male) presenting with a CIS were identified from the total cohort (Chapter 3.1.1). Subjects had clinical and MRI assessment at baseline, three months and one year, and clinical follow-up at five years. Baseline assessment was performed within 12 weeks of first presentation (median 5.5 weeks) and mean age was 34.6 years (SD 6.6). Initial presentation was optic neuritis in 34 subjects, spinal cord syndrome in two subjects and brainstem syndrome in one subject. Subjects were assessed for progression to MS according to the McDonald criteria (McDonald *et al*., 2001). Coronal three-dimensional inversion recovery prepared FSPGR, and FSE sequences, were acquired for all subjects at baseline using the acquisition parameters described in Chapter 3.3.1. A coronal FSPGR image acquired again at one year was also used for this analysis.

MRI analysis

T2-weighted images were assessed for the presence of lesions at baseline, and the number of lesions were determined by a trained neurologist (K. Fernando). Brain atrophy was calculated from the FSPGR images using SIENA, as described in Chapter 5.3.2., and results were annualised.

Statistical analysis

Comparisons of demographic data, mean brain atrophy rates and proportion of subjects with T2 lesions at baseline was performed between groups (subjects who had progressed to MS by five year follow-up (“converters”) and those that had not (“non-converters”))

using two-tailed unpaired *t*-tests and Fisher's exact test where appropriate. A logistic regression model was used to relate five year clinical status jointly to the presence of lesions at presentation and early brain atrophy rate. Odds ratios (OR) and 95% CI were calculated to determine the predictive value of T2 lesions at presentation and early (baseline to one year) brain atrophy rate for progression to MS. However, a difference between the p-values for the Wald and likelihood ratio (LR) tests suggested that the logistic regression model results should not be completely relied on. Therefore, to assess the extent to which initial brain atrophy rate was independently useful in differentiating converters and non-converters, a linear regression model was used to relate brain atrophy rate to five year clinical status whilst controlling for lesions at presentation. Robust standard errors were used in these models to allow for differential heterogeneity between subjects in each of the two groups. Survival analysis was used to investigate the predictive value of brain atrophy rate and T2 lesions at baseline for the risk of subsequent MS diagnosis. A Cox multivariable regression analysis was used to relate the risk of MS diagnosis to early brain atrophy rate, presence of T2 lesions on baseline MR, and the number of T2 lesions detected on baseline MRI. Age at follow-up and gender were also included as covariates. The influence of atrophy rate on time to progression was further investigated with Kaplan-Meier survival curves. Subjects were divided into three groups of approximately equal size and according to their atrophy rates relative to the control brain atrophy rates determined in Chapter 6.2; group A (n= 14) had atrophy rates less than or equal to the control rate (i.e. $x \geq -0.07$), group B (n=11) had atrophy rates greater than the control rate but less than one SD from the control rate (i.e. $-0.41 \leq x < -0.07$), and group C (n=12) had atrophy rates greater than one SD from the control rate (i.e. $x < -0.41$). These cut-off points were chosen because they divided the subjects into groups of similar size. A plot of the Kaplan-Meier estimates of survival function for the three groups was generated and differences between the three groups were analysed by the log-rank test, taking into account the fact that the groups were ordered.

9.2.3 Results

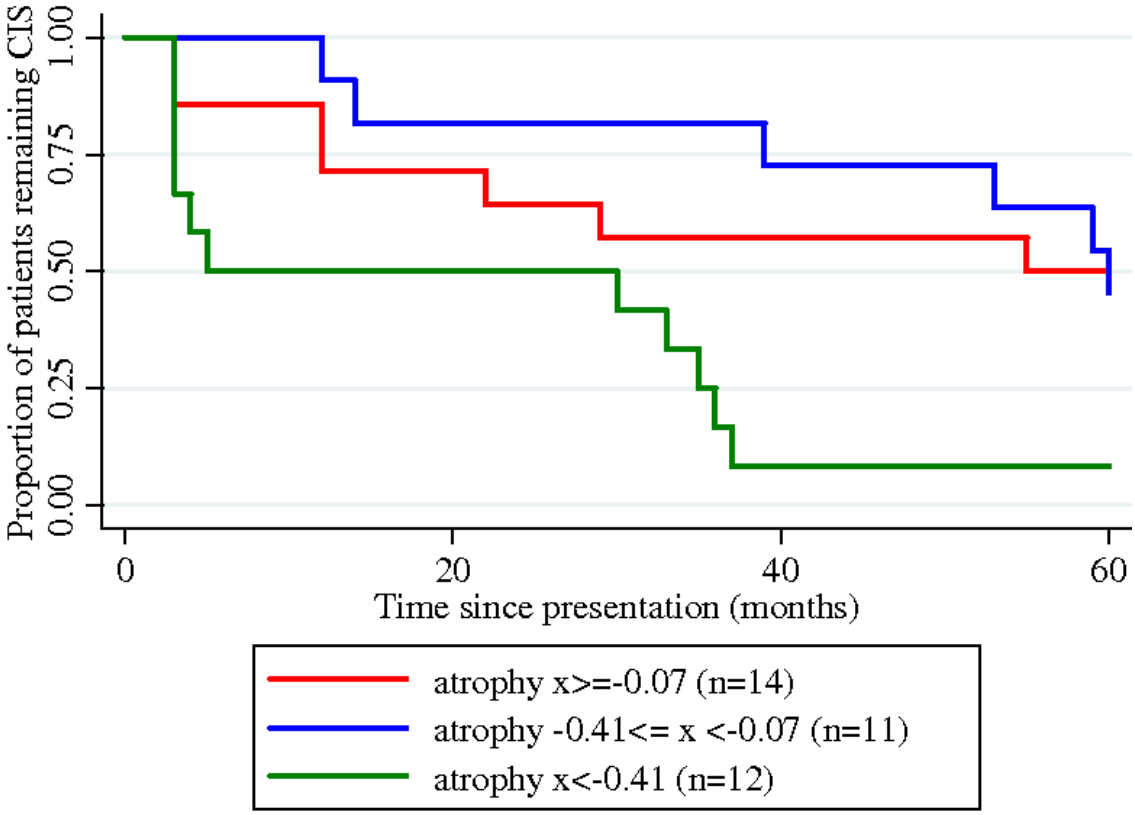
At five year follow-up 24 subjects (seven male) had progressed to MS according to the McDonald criteria, whilst 13 subjects (seven male) remained CIS. There were no significant differences in age (p=0.78), gender (p=0.17), time since presentation (p=0.73) or interval between baseline and five year follow-up (p=0.52), between converters and non-converters. Mean atrophy rate in non-converters was 0.02% $year^{-1}$ (SD 0.29), whilst in converters it was -0.44% $year^{-1}$ (SD 0.49). There was evidence that atrophy rates in the

two groups were significantly different (mean difference 0.46% year^{-1} (95% CI 0.16 to 0.77, p=0.001)). Of the 24 converters, 23 had T2 lesions (any number, i.e. ≥ 1) on MRI at presentation whilst of the 13 non-converters, five demonstrated T2 lesions on MRI at presentation (Fisher's exact test p<0.001).

Univariate logistic regression analysis provided no evidence that gender, age or the number of lesions at baseline were significant predictors of a subsequent diagnosis of MS, and these variables were therefore excluded from the multiple regression analysis. Multiple logistic regression showed that the odds of progressing to MS were approximately 25 times greater if subjects presented with T2 lesions (95% CI 2 to 287 times, Wald test p=0.009, LR test p=0.002), and approximately 22 times greater per 1% year^{-1} increase in brain atrophy rate (95% CI 0.7 to 668 times, Wald test p=0.074, LR test p=0.027). These effects were similar in magnitude to results from models assessing the effect of each variable separately. These results suggest that atrophy rate may be an independent predictor of MS diagnosis, but the difference between the p-values for the Wald and LR tests suggest that the logistic regression model should not be completely relied on. Using linear regression with robust standard errors, atrophy rate differed significantly between converters and non-converters even after adjustment for the presence of lesions at baseline (p=0.006), providing reliable evidence that atrophy rate is independently predictive of five year clinical status.

In the Cox regression model it was found that the presence of lesions at baseline and atrophy rate during the first year were risk factors for MS diagnosis. The risk (hazard) of diagnosis with MS was multiplied by 3.6 (95% CI 1.3 to 10.0, p=0.016) for each 1% year^{-1} increase in atrophy rate, and by 11.5 (95% CI 1.4 to 91.7, p=0.021) when lesions were present at baseline. The number of T2 lesions at baseline, age and gender were not independent risk factors for MS diagnosis. When subjects were divided into three groups according to one year atrophy rate, there was evidence that subjects with greater atrophy rates during the first year were at a greater risk of subsequently being diagnosed with MS; median survival times for groups A, B and C respectively were 57.5 months, 60 months and 17.5 months, p=0.015 (Figure 9-1).

Figure 9-1 *Kaplan-Meier curves for subjects presenting with CIS according to brain atrophy rate (% year^{-1}) during the first year after presentation.*



9.2.4 Discussion

This study provides evidence that brain atrophy occurring early in the course of disease is clinically relevant. Brain atrophy rate during the first year following presentation was found to be a significant predictor of a future diagnosis of MS, which was independent of the predictive value of T2 lesions at presentation. Evidence was also found for a significant relationship between brain atrophy rates during early follow-up and the risk of being diagnosed with MS, with greater atrophy rates associated with a higher risk.

The finding that brain atrophy occurs from the earliest stages of disease and that mean brain atrophy rate is greater in subjects who progress to MS than in subjects who remain stable has been shown previously (Brex *et al.*, 2000; Dalton *et al.*, 2002a; Dalton *et al.*, 2004). Although not investigated in this study, it has been suggested that this atrophy is related to focal inflammatory lesions (Dalton *et al.*, 2002a; Filippi *et al.*, 2004; Paolillo *et al.*, 2004). However lesions do not wholly account for the variance in atrophy rate (Dalton *et al.*, 2002a; Paolillo *et al.*, 2004) and other, perhaps unknown, mechanisms for neuroaxonal damage may be occurring at this stage of the disease.

In contrast to other studies of brain atrophy in CIS, subjects in this study had a longer clinical follow-up and the value of brain atrophy for future prognosis was investigated. The previously recognised value of T2 lesions on MRI at presentation for predicting progression to MS (Brex *et al*., 2002; Ghezzi *et al*., 1999; Optic Neuritis Study Group, 1997) was confirmed in this study. Logistic regression estimated that subjects were 25 times more likely to progress to MS if T2 lesions were present on baseline MRI. Due to the dichotomous nature of this predictor variable it was not possible to verify the result using linear regression, but it was highly significant when tested by both the Wald and LR tests. Unlike other studies, the number of T2 lesions at baseline was not prognostic (Brex *et al*., 2002; Tintoré *et al*., 2006). Regression analysis also showed that brain atrophy rate during the first year after presentation was a significant predictor of progression to MS, independent of T2 lesions on baseline MRI. This finding was in contrast to a study of 35 patients with CIS, who showed evidence of disease dissemination in space (Rovaris *et al*., 2005b). Despite 24 patients progressing to MS at the end of the one year study, brain atrophy measured by SIENA over the first year from presentation did not significantly predict subsequent diagnosis. This may have been due to differences in the cohorts; only 37% of subjects presented with optic neuritis in contrast to 92% of subjects in this study. Indeed one of the limiting factors of this study is that the majority of subjects presented with optic neuritis. Any association between presenting symptoms and subsequent progression could not be investigated. Although optic neuritis has been associated with a lower risk for conversion to MS, this appears to be related to normal MRI being more commonly observed in these patients (Tintoré *et al*., 2005). Risk of MS conversion was similar in patients with different presenting symptoms when only those subjects with abnormal MRI were studied. Another study observed that in patients with an abnormal baseline scan, progression to MS is highest in those presenting with optic neuritis, followed by those with a brainstem syndrome, and the lowest rates were seen in patients with spinal cord syndrome (Morrissey *et al*., 1993). This may account for the discrepancy of findings between this study and that by Rovaris *et al.* (Rovaris *et al*., 2005b).

Evidence from this study that progressive brain atrophy is greater in patients who progress to MS earlier, supports the hypothesis that even in CIS patients the more active the disease the greater the accumulation of permanent damage. The number of lesions (T2 and Gd-enhancing) on baseline MRI has been associated with an early MS diagnosis (Pestalozza *et al*., 2005), and patients with worse clinical outcomes have been shown to have larger numbers and volumes of lesions on MRI at baseline (Brex *et al*., 2002). In

addition, greater increases in T2 lesion volumes have been seen in subjects who develop MS compared with those that remain CIS (Dalton *et al*., 2004), and in patients with abnormal MRI at baseline a trend has been shown for those with more lesions to develop MS more frequently (Morrissey *et al*., 1993). However the finding that higher atrophy rates were associated with a higher risk of MS diagnosis was most likely driven by the patients with particularly high atrophy rates. In fact, the Kaplan-Meier plot shows that survival without progression to MS was very similar for patients with intermediate atrophy rates and for patients with atrophy rates at the level of controls or less during the first year. This may have been due to the cut-offs used to group patients which, although based on mean control atrophy rates and dividing the cohort equally, are relatively arbitrary. In fact the 95% CI for the atrophy rate in control subjects was -0.25 to 0.11, which spans atrophy rates included in the intermediate atrophy group (B), suggesting that a significant proportion of CIS patients have cerebral losses over one year that are within the control range. Those subjects with rates outside this range (group C) have much higher rates of conversion to MS at five years. It is also likely that pathology occurring after the one year observation period in subjects from groups A and B will influence whether patients develop MS. This illustrates that although brain atrophy rate may be a significant independent risk factor for progression to MS, it would be difficult to incorporate this indicator for diagnostic purposes or development of management strategies. There is difficulty determining what is normal and abnormal, and a continuum of atrophy rates exists in CIS subjects that may also vary depending on the measurement technique, person measuring the atrophy and the scanner and acquisition. In addition, pathology occurring beyond one year follow-up may contribute to progression to MS.

Further investigation of the prognostic value of brain atrophy rate in larger cohorts of CIS subjects with longer clinical and MRI follow-up need to be performed. In addition to the small number of subjects included in this study, the investigation presented has several limitations, some of which have already been mentioned. Partly as a result of the small sample size, it was not possible to include a large number of explanatory variables in relation to subsequent MS diagnosis. Larger, prospective studies would have the power to investigate additional MRI, clinical, demographic and genetic factors that may aid in the prognosis for patients presenting with a CIS. Also, the log-rank test showing a significant relationship between atrophy rate and the survival time does not provide a comparison of the total survival experience of the three groups, but rather gives a comparison at an arbitrary time-point (in this study the median survival time). In addition, the survival

analysis included subjects who had been diagnosed with MS during the first year of follow-up, when measurement of brain atrophy had been performed. However it is impossible to measure brain atrophy rate prior to clinical presentation, and excluding those patients who converted during the first year is likely to bias the results. Therefore as this experiment was exploratory it was felt that all patients should be included.

Although clinical follow-up was extended to five years in this study, the time during which there is the highest probability of developing MS, of the 13 subjects remaining clinically isolated at follow-up, some could still progress to MS. In a 14 year follow-up study of 71 CIS patients, amongst those with a normal baseline MRI scan who converted to clinically definite MS, the median time to development of MS was 7.5 years (range 5 to 11 years) (Brex *et al*., 2002). However that study found that approximately 68% of patients overall had developed MS at follow-up, which is a similar proportion to that found at follow-up in this study (65%). In contrast, a 10 year follow-up study of 102 subjects presenting with optic neuritis found that the risk of progressing to MS was only 42% after 10 years (Ghezzi *et al*., 1999). Obviously it cannot be concluded with any certainty that of the patients remaining stable at five year follow-up in this study few will progress to MS at a later date, as the cohorts are different and MS is a heterogeneous disease, but these studies might be considered suggestive that the majority of patients in this study who are going to progress to MS have already done so.

In conclusion, the findings in this investigation indicate that initial brain atrophy rate following presentation with a CIS could be an additional prognostic factor in patients presenting with a CIS. These results need to be confirmed in bigger cohorts with longer clinical and MRI follow-up.

9.3 Investigation into the predictive value of brain atrophy rate for clinical disability in patients with relapsing remitting multiple sclerosis

9.3.1 Introduction

Many studies previously investigating the association between brain atrophy and subsequent disability may have been limited by the cross-sectional study design (Calabrese *et al*., 2007a; Chard *et al*., 2002b; Filippi *et al*., 1998; Ge *et al*., 2001; Quarantelli *et al*., 2003), or short clinical follow-up which could have been insufficient to detect a significant change in disability using EDSS scores (Gasperini *et al*., 2002; Luks *et al*., 2000). This may explain the weak or absent associations observed between atrophy

and disability in RRMS (De Stefano *et al*., 2003; Lin *et al*., 2003; Losseff *et al*., 1996; Luks *et al*., 2000; Lycklama á Nijeholt *et al*., 1998; Rovaris *et al*., 2003). Associations between changes in global or regional brain volumes and concurrent changes in EDSS score have also been mixed (Audoin *et al*., 2006; Audoin *et al*., 2007c; Kalkers *et al*., 2002; Rudick *et al*., 2000). Several reasons have been suggested for these findings, including that brain atrophy may not immediately lead to changes in disability, that there may be a threshold of neuroaxonal damage before which disability is not clinically apparent, that cortical reorganisation may limit clinical manifestations, and that neuroaxonal damage may occur in clinically silent areas. It has also been suggested that the EDSS may be too heavily weighted towards motor disability, it is insensitive to small amounts of change and is subject to high inter-rater variability (Noseworthy, 1994).

Longer clinical follow-up studies have greater potential to determine relevant associations between brain atrophy and subsequent development of disability. Inclusion of longitudinal measures of brain atrophy, as opposed to cross-sectional volumes, may allow additional associations to be seen. Furthermore, new MS rating scales have been developed, including the MSFC (Cutter *et al*., 1999), which is thought to provide a more sensitive measure of clinical disease progression than the EDSS (Cutter *et al*., 1999; Hobart *et al*., 2004). A study in 2001showed MRI measures of brain T1 and T2 lesion loads to correlate significantly with MSFC scores but not EDSS scores in a large group of relapse-onset patient (Kalkers *et al*., 2001b). The MSFC is comprised of three objective and quantitative tests: leg function/ambulation, arm/hand function, and cognitive function, and standardised z-scores are used to relate an individual's performance to the average performance in the population.

The aim of this study was to investigate the relationship of annual atrophy rate early in the course of RRMS to the development of disability by five year follow-up, as measured by the MSFC, and investigate whether brain atrophy might explain subsequent clinical disability better than established MRI measures of focal inflammation.

9.3.2 Methods

Subjects and MR imaging

Twenty-five subjects (six males) with RRMS were identified from the total cohort of 41 subjects enrolled in a longitudinal clinical and MRI study (Chapter 3.1.2). All subjects had been recruited within four years of symptom onset (median 1.7, range 0.5-3.8), and

had been assessed at baseline, one year and five years. Median EDSS at baseline was 1.5 (range 0-2). Mean follow-up time was 5.0 years (SD 0.92) when mean age was 42.1 years (SD 7.5) and mean disease duration was 6.9 years (SD 1.3). Seventeen patients were on disease-modifying treatments at follow-up. At baseline a coronal T1-weighted three-dimensional inversion recovery prepared FSPGR, a T1-weighted CSE and a FSE sequence were acquired using the sequences outlined in Chapter 3.3.2. The T1-weighted three-dimensional inversion recovery prepared FSPGR was repeated at approximately one year follow-up.

MRI analysis

T1- and T2-weighted lesion areas were identified on the T1-weighted CSE and PD-weighted images respectively, and contoured using DispImage (Plummer, 1992) (Chapter 3.4.4) by a trained neurologist (W. Rashid). Annual percentage brain atrophy rate was determined from the baseline and one year T1-weighted FSPGR images using SIENA, according to the optimised method described in Chapter 5.3.2.

Multiple Sclerosis Functional Composite score

The MSFC was administered at baseline and five year follow-up as outlined in Appendix 2 (Cutter *et al*., 1999; Fischer *et al*., 1999). Z-scores at follow-up were calculated using the baseline scores from the whole RRMS cohort (n=41) as the reference population.

Statistical analysis

Univariate linear regression was performed to investigate linearity between each independent variable (MRI and demographic factors) and the dependent variable (MSFC z-score). Multiple linear regression analysis was performed to determine significant relationships between MSFC z-score and MRI or demographic variables, using a forward stepwise procedure. A p-value of 0.05 was required for a variable to be included in the model and p-value of 0.1 was required to retain the variable in the final model. Predictor variables investigated were baseline T1 lesion volume, baseline T2 lesion volume, baseline to one year atrophy rate, disease duration, age at follow-up, interval (baseline to five year follow-up), gender and treatment status at five year follow-up.

9.3.3 Results

Mean atrophy rate in the 25 subjects was -0.81% $year^{-1}$ (SD 0.50). Mean baseline T1-weighted lesion load was 1.30cm^3 (SD 1.28, range 0-3.75) and mean baseline T2-

weighted lesion load was 6.84cm^3 (SD 4.64, range 0.86-19.61). Median EDSS at five year follow-up was 2.5 (range 0-6) whilst mean MSFC z-score was -0.10 (SD 1.10, range -3.13 to 1.16). Using forward stepwise linear regression, brain atrophy rate during the first year of study, treatment status at follow-up and baseline T1 lesion volume were all retained in the final model as significant independent predictors of MSFC z-score at five years (Table 9-1). T2 lesion volume, disease duration, age at follow-up, interval and gender were found not to be significant independent predictors of MSFC z-score at five year follow-up (Table 9-1). For each 1% year^{-1} increase in atrophy rate, MSFC z-score decreased by 1.03 (95% CI 0.38 to 1.69, p=0.003). Similarly, when on treatment MSFC z-score decreased by 1.01 (95% CI 0.31 to 1.70, p=0.007). For each 1cm^3 increment in baseline T1-weighted lesion volume, MSFC z-score decreased by 0.29 (95% CI 0.031 to 0.54, p=0.030). The final model accounted for 58% of the variance in MSFC z-score.

Table 9-1 *Crude and adjusted regression coefficients for independent variables entered into a linear regression analysis looking at association with MSFC z-score.*

Predictor variable	***Effect on MSFC z-score***	
	Univariate regression coefficient (β), (95% CI)	***Adjusted regression coefficient (β), (95% CI)***
Baseline T1 lesion volume (cm^3)	-0.32 (-0.66, 0.023) p=0.066	-0.29 (-0.54, -0.031) **p=0.030**
Baseline T2 lesion volume (cm^3)	-0.033 (-0.13, 0.068) p=0.502	0.084 (-0.095, 0.26) p=0.339
Brain atrophy rate (% year^{-1})	1.07 (0.24, 1.90), p=0.014	1.03 (0.38, 1.69) **p=0.003**
Disease duration (years)	0.24 (-0.13, 0.60) p=0.189	0.12 (-0.16, 0.41) p=0.377
Age at follow-up (years)	-0.0031 (-0.067, 0.061) p=0.921	0.017 (-0.031, 0.034) p=0.474
Interval baseline to 5 years (years)	0.37 (-0.12, 0.86) p=0.133	0.28 (-0.093, 0.65) p=0.134
Gender (male vs female)	0.29 (-0.79, 1.38) p=0.579	-0.41 (-1.21, 0.39) p=0.300
Treatment status (no DMT vs DMT)	-1.19 (-2.05, -0.34) p=0.008	-1.01 (-1.70, -0.31) **p=0.007**

9.3.4 Discussion

This investigation has provided evidence that brain atrophy rate early in the course of MS (within four years of symptom onset) is significantly related to subsequent disability measured by the MSFC. In addition, the volume of T1-weighted lesions, which are thought to be the result of severe focal tissue damage, was also a significant independent predictor of disability five years later. These findings are important in the search for valid markers of disease progression in MS and understanding disease mechanisms.

One of the strengths of this study was that patients had a relatively long clinical follow-up, and had been recruited into the study within four years of symptom onset. Partly because of the mixed correlations between atrophy and disability over shorter follow-up periods, it is thought that effects on disability may lag relative to changes in rates of brain atrophy. Studies such as this one, which include longer patient follow-up may aid in establishing the relationship between brain atrophy and permanent disability. One such long-term follow-up study of 138 RRMS patients who had taken part in a treatment trial of beta interferon, found that brain atrophy during the original two year trial was significantly correlated to MSFC at eight year follow-up (r=0.35). Moreover, in a logistic regression analysis, brain atrophy during the original two year trial was a significant predictor of patients having an EDSS greater than or equal to six (Fisher *et al*., 2002). However the patients had longer disease duration and larger lesion volumes at baseline than the cohort investigated in the current study, which suggests that even an increased atrophy rate early in the course of the disease may be relevant to future disability. A recent study published in 2007, investigating patients with a disease duration of less than two years, found that atrophy rate over approximately two years was significantly related to MSFC score at follow-up (Jasperse *et al*., 2007c).

The findings in a six year follow-up study of 55 RRMS patients who had been enrolled in a two year trial of interferon beta-1a were similar to this study, in that T1 lesion volume at baseline was a significant predictor of the change in disability over the six years (Paolillo *et al*., 2002). However in contrast to the current study, T2 lesion volume at baseline was also a significant predictor of the change in disability whilst atrophy was not. That study utilised the EDSS as a measure of disability, which as already mentioned may be subject to interater variability, and therefore may to some extent explain the difference in findings. The difference may also be because brain atrophy was measured using the CCV which, as has been shown in Chapter 6.3, may not be as sensitive or precise in

measurements as SIENA. Inconsistencies in the association of atrophy with disability due to differences in the atrophy measurement technique have been seen in other studies. A four year investigation of 20 patients with RRMS and 18 patients with SPMS showed no significant difference in the change in EDSS over this period between subjects classified with brain atrophy and those without brain atrophy (Turner *et al.*, 2003). However significantly greater disability was observed in the same patients classified as having ventricular enlargement compared with those who did not – this discrepancy suggests a difference in the sensitivity of methods.

In contrast to the current study, several investigations have shown that T2 lesion volume correlates with or is predictive of EDSS. In 142 RRMS patients taking part in a trial of glatiramer acetate, a correlation was observed between T2 lesion volume at baseline and EDSS approximately six years later (Rovaris *et al*., 2007). A smaller study that included patients with both RRMS and SPMS found baseline T2 lesion volume was predictive of EDSS deterioration at a follow-up of approximately 4.5 years (logistic regression analysis of stable versus sustained increase in EDSS) (Rovaris *et al*., 2003). Baseline T1 lesion volume and atrophy rate measured over the first year of the study using SIENA were not significantly associated. An eight year follow-up of the same patient cohort confirmed baseline T2 lesion volume was a predictor of EDSS worsening, and by this stage T1 lesion volume was also a significant independent predictor (Agosta *et al*., 2006). That baseline T1 lesion volume became predictive of EDSS worsening at a later stage may be further evidence of a delay between permanent tissue damage and disability. Although brain atrophy remained a non-significant predictor of EDSS worsening in the eight year follow-up study, analysis included changes in BPF, GMF and WMF. Again, the discrepancy in results with the current study could be because different atrophy measurement techniques were used and the fact that EDSS was utilised as the measure of clinical disability.

Interestingly, it was observed in this analysis that patients who were on disease-modifying treatments had significantly worse disability measured by the MSFC at five year follow-up. This is likely to be because patients who had a more aggressive disease course were placed on treatment, rather than that treatment was causing increased disability.

Limitations of this study should be noted. Firstly, although the MSFC is thought to be a more sensitive marker of clinical disability than other clinical scales, including the EDSS,

it may still be subject to error and it is difficult to determine the actual amount of permanent disability that a patient has accumulated. One of the primary reasons for investigating markers of disease progression is that pathology may be occurring in clinically silent locations, and therefore rating scales and MRI markers measure different manifestations of MS and are unlikely to correlate completely. A further limitation of the study is that the MSFC had been administered in these patients at approximately six-monthly intervals for three years prior to the five year follow-up assessment. Practice effects, particularly with regard to the PASAT may therefore have influenced the MSFC score to some extent. Also, it was not possible to investigate other clinical and MRI variables that may be related to subsequent disability, such as relapse rate and quantitative measures from DW-MRI. This was in part due to the small sample size which did not allow the inclusion of many more variables in the regression model. This study did not investigate concurrent changes in atrophy and disability either, which has been presented in some studies previously (Fisher *et al*., 2000; Rudick *et al*., 2001), and may help to further elucidate the relationship between brain atrophy and disability. Investigation of associations between regional brain atrophy rates and scores of localised function may also aid in this endeavour (Calabrese *et al*., 2007a).

In conclusion, these results provide further evidence that neuroaxonal damage is a cause of permanent disability in MS which may precede clinical manifestations, and that brain atrophy may be a good marker of that damage and thereby of disease progression. MRI measures of disease activity in the early years of the disease appear to be important in the long-term prognosis for disability in MS patients. Larger studies and investigation of the associations of atrophy and disability in progressive MS need to be performed.

9.4 Investigation into the predictive value of brain atrophy rate for cognitive impairment in patients with relapsing remitting multiple sclerosis

9.4.1 Introduction

As described in Chapter 2.6.2, cognitive impairment is known to occur in patients with MS. It is thought that between approximately 40-60% of MS patients exhibit cognitive impairment (Achiron & Barak, 2003; Rao *et al*., 1991), and subtle cognitive changes have been detected in subjects presenting with a CIS (Feuillet *et al*., 2007) and subjects with early RRMS (Amato *et al*., 2001; Deloire *et al*., 2005). Deficits in attention, speed of information processing, working memory and verbal and visuospatial memory are commonly observed (Amato *et al*., 2001; Benedict *et al*., 2005; Benedict *et al*., 2006;

Deloire *et al*., 2005; Lazeron *et al*., 2006). These deficits may significantly impact on daily living in relation to both occupational and social functioning. Cognitive impairment is progressive over the course of the disease and deficits may extend to additional domains and functions over time (Amato *et al*., 2001).

Associations between global and regional brain volumes and cognitive dysfunction have been shown in cross-sectional studies (Amato *et al*., 2004; Christodoulou *et al*., 2003; Sanfilipo *et al*., 2006). Moreover brain atrophy has been shown to account for greater variance in cognitive functioning than lesion burden in cross-sectional studies (Benedict *et al*., 2004; Lazeron *et al*., 2006). However some cross-sectional studies have not shown significant associations between cognitive impairment and MRI measures of disease burden (Achiron & Barak, 2003; Deloire *et al*., 2005). The true relationship of brain atrophy and cognitive changes need to be explored further through longitudinal studies. Little is known about the ability of MRI parameters, including early brain atrophy, to predict the development of specific cognitive deficits. Identification of patients who might go on to develop such deficits may aid in the screening of patients and provide a rationale for administration of disease-modifying therapy, whilst in addition allowing management strategies to be developed, and support for individual patients. This study explores the value of MRI parameters obtained at baseline in predicting cognitive impairment in specific domains five years later.

9.4.2 Methods

Subjects and MR imaging

Of the initial 41 patients with RRMS (Chapter 3.1.2), 34 subjects had returned for cognitive assessment at five year follow-up. Mann-Whitney tests showed no significant differences between subjects that returned for cognitive assessment and those that did not in age at onset, disease duration, EDSS at five year follow-up, baseline lesion volume, or brain atrophy rate during the first year of follow-up. Of the subjects that returned for cognitive assessment, 26 (six male) had baseline and one year follow-up MRI scans available for atrophy measurement. Subjects had been recruited within four years of first presentation. At baseline, mean age was 37.2 years (SD 7.5), mean disease duration was 2.0 years (SD 0.8) and median EDSS was 1.5 (range 0-3). At baseline and one year follow-up assessment a coronal T1-weighted three-dimensional inversion recovery prepared FSPGR, a T1-weighted CSE and a FSE sequence were acquired using the parameters outlined in Chapter 3.3.2.

MRI analysis

Annual percentage brain atrophy rate over the first year from baseline assessment was determined using SIENA on T1-weighted coronal FSPGR images, according to the optimised method described in Chapter 5.3.2. T1- and T2-weighted lesion areas were identified on the baseline T1-weighted CSE and PD-weighted images respectively and contoured using Dispimage (Plummer, 1992) (Chapter 3.4.4) by a trained neurologist (W. Rashid).

Clinical and neuropsychological assessment

At five-year follow-up, self-reported symptoms of anxiety and depression were rated with the Hospital Anxiety and Depression Scale (HADS) (Snaith & Zigmond, 1986). A score of 11 or above was used as a cut-off point for both subscales.

A neuropsychological test battery was administered at five year follow-up by a trained psychologist (M. Summers) and the following cognitive domains were assessed:

(i) General intellectual functioning was assessed with a shortened version of the Wechsler Adult Intelligence Scale – Revised (WAIS-R) (Wechsler, 1981) and measures of verbal (vocabulary, digit span, arithmetic and similarities subtests), performance (picture completion, picture arrangement and block design subtests) and full-scale IQ were obtained. Optimal premorbid intellectual functioning was estimated using the National Adult Reading Test (NART) (Nelson, 1982). The difference between the NART IQ and the WAIS IQ was considered an index of IQ change. Patients with measures of current IQ 15 or more points below premorbid estimates were considered to have IQ decline.

(ii) Verbal and visual recall memory were assessed using the story and figure recall subtests of the Adult Memory and Information Processing Battery (AMIPB) (Coughlan & Hollows, 1985). In the story recall subtest, participants are asked to freely recall a short story immediately after presentation and again after a 30-minute delay. The score is the number of correctly recalled story segments, out of a maximum 56. In the figure recall subtest, participants must copy a complex figure, then re-draw it from memory immediately, and after a 30-minute delay. The score is the number of correctly drawn design elements, out of a maximum 80.

(iii) Attention/speed of information processing were assessed with the PASAT (Sampson, 1956) and the Symbol Digit Modalities Test (SDMT) (Smith, 1982). In the PASAT, subjects have to add successive pairs of digits which are presented aurally at three second intervals. The score is the number of correct additions, out of a maximum 60. In the SDMT, subjects have to transcribe single digits from symbols according to a visually presented key. The score is the number of digits correctly transcribed in 90 seconds.

(iv) Executive functioning was assessed using the Spatial Working Memory (SWM) subtest of the Cambridge Neuropsychological Test Automated Battery (CANTAB) (Sahakian & Owen, 1992), the Hayling Sentence Completion Test (Burgess & Shallice, 1997) and the Brixton Spatial Anticipation Test (Burgess & Shallice, 1997). In the SWM subtest, participants search for tokens in an array of boxes (four-, six- and eight-box arrays) presented on a computer monitor, remembering and avoiding locations where they have already found a token. The number of times participants return to a location where a token has been found during an earlier search is the error score. A strategy score is also given, which is a measure of search pattern consistency, with a lower strategy score indicating a more efficient search strategy. The Hayling sentence completion task consists of two parts. Part one measures verbal response generation, where participants must complete a sentence with an appropriate word, and part two measures response suppression, where participants must complete a sentence with an unconnected word, suppressing connected response words. The score is derived from a composite of response times for each item in parts one and two and the number of errors (connected word responses) made in part two. Longer response times and more errors yield a lower score. In the Brixton test, participants are presented with a 10-position array, in which one position is marked by a filled circle. On subsequent presentations, the marked position changes according to a pattern and participants are required to guess which position will be marked on the subsequent presentation, based on the current spatial pattern of movement. The score is the total number of incorrect guesses made.

Analysis of neuropsychological data

To determine whether patients were impaired in each cognitive domain, patients' raw scores were compared with those from age-related healthy controls in the published literature, and converted into percentiles. As the normative controls came from a variety of educational backgrounds, cognitive impairment was not adjusted for educational level. Following standard practice, scores falling at or below the 5th percentile of published

norms were considered impaired. The total number of tests failed (i.e. performance at or below the 5th percentile of published norms) was also determined for each patient.

Composite and domain-specific z-scores were obtained in order to analyse associations between MRI parameters and neuropsychological performance. For each test score, mean and standard deviation was calculated for the 26 patients, allowing each patient's score to be expressed as a z-score, referenced against other patients' performance. Z-scores from measures of IQ deficit were averaged to produce an IQ deficit z-score (such that a smaller IQ deficit constituted a higher z-score). Likewise, z-scores from all memory, attention and executive function tests were averaged to produce memory, attention and executive function z-scores respectively. These z-scores were averaged for each patient to produce a composite z-score in which each cognitive domain was equally weighted.

Statistical analysis

Multiple linear regression analysis was performed to determine significant relationships between cognitive performance and MRI variables. Due to the small sample size in this study and the large number of demographic and clinical variables, univariate linear regression analyses were performed for each variable and only those that were significant were retained for the multivariable model. The variables considered were disease duration, age at follow-up, interval (baseline to five year follow-up), gender, medication, years of education, EDSS at follow-up, premorbid IQ and anxiety and depression ratings. Significant variables were retained and entered into the multivariable linear regression using a forward stepwise procedure with the MRI predictor variables which included baseline to one year brain atrophy rate, and baseline T1 and T2 lesion volumes. A p-value of 0.05 was required for a variable to be included in the model and p-value of 0.1 was required to retain the variable in the final model.

9.4.3 Results

At five year follow-up mean age was 42.2 years (SD 7.5), mean disease duration was 7.0 years (SD 1.2) and median EDSS (based on only 24 subjects) was 2.5 (range 0-6). Fifteen patients were on disease-modifying treatment and the mean interval between baseline and follow-up assessments was 5.0 years (SD 0.9). Two patients were not rated with the HADS, whilst three patients did not complete the NART and consequently IQ deficit could not be established for these subjects. At follow-up, the HADS scores were above the threshold for depression in one patient and above the threshold for anxiety in six.

In the 23 subjects in whom premorbid IQ was established, seven (30%) showed IQ decline from premorbid estimates in at least one IQ measure (five verbal IQ, five performance IQ, four full-scale IQ). Scores for individual neuropsychological tests and the number of subjects impaired at five year follow-up are shown in Table 9-2.

Table 9-2 *Standardised neuropsychological test scores and numbers of patients impaired at five year follow-up.*

Test		***Mean (SD)***	***Range***	***Impaired***	***Composite z-score (mean)***
General intelligence	*Premorbid full-scale IQ*	108.6 (10.1)	86-124	0/23	2.11×10^{-9}
	Full-scale IQ deficit	8.8 (6.8)	-5-22	4/23	
Memory	*Story recall: immediate (max 56)*	36.6 (8.2)	19-53	0/26	
	Story recall: delayed (max 56)	34.1 (9.5)	18-53	0/26	
	Figure copy (max 80)	78.1 (2.5)	70-80	0/26	-4.01×10^{-9}
	Figure recall: immediate (max 80)	61.2 (14.3)	28-80	1/26	
	Figure recall: delayed (max 80)	60.0 (13.2)	38-80	0/26	
Executive function	*SWM: within-trials error*	3.5 (4.8)	0-16	2/26	
	SWM: strategy	33.2 (7.2)	8-42	0/25	
	Hayling test: overall score (max 23)	17.4 (3.1)	6-21	0/25	2.31×10^{-3}
	Brixton test: errors (max 54)	14.4 (6.1)	3-29	2/25	
Attention	*PASAT 3 (max 60)*	46.5 (15.0)	0-60	3/26	1.50×10^{-9}
	SDMT	46.4 (10.8)	23-65	6/26	
Overall cognition					-0.021

PASAT, Paced Auditory Serial Addition Test; SDMT, Symbol Digit Modalities Test; SWM, Spatial Working Memory.

Mean atrophy rate in the 26 subjects was -0.81% year^{-1} (SD 0.50). Mean baseline T1-weighted lesion load was 1.51cm^3 (SD 1.63, range 0-6.62) and mean baseline T2-weighted lesion load was 7.34cm^3 (SD 5.40, range 1.42-21.3).

Of the demographic and clinical variables, gender was associated with IQ deficit, premorbid IQ was associated with tests of attention and executive function, and anxiety rating was associated with executive function. Using multiple regression, brain atrophy rate in the first year from baseline was retained in the final models as the only significant independent predictor for memory z-score and overall cognitive z-score. For each 1% year^{-1} increase in atrophy rate, memory z-score decreased by 0.65 (95% CI 0.09 to 1.20, p=0.024) and overall cognitive z-score decreased by 0.56 (95% CI 0.15 to 0.96, p=0.009). Brain atrophy rate accounted for 16% and 22% of the variance in memory and cognitive z-scores respectively. Baseline T1 lesion volume was the only significant predictor retained in the model for attention z-score. For each 1cm^3 increase in T1-weighted lesion volume, attention z-score decreased by 0.21 (95% CI 0.03 to 0.38, p=0.021). IQ deficit and executive function were not significantly predicted by atrophy rate or lesion volumes.

9.4.4 Discussion

This study has demonstrated that a higher brain atrophy rate early in the course of the disease is associated with lower scores on tests of cognition in patients with RRMS. Specifically, a higher brain atrophy rate was found to be an independent predictor of overall cognitive functioning and memory impairment. In addition, T1 lesions were found to be an independent predictor of performance in tasks of attention. These results were found even when clinical and demographic variables had been taken into account and are unlikely to have been influenced by the education level of subjects, as years of education and premorbid IQ were included as variables.

At five year follow-up the patients included in this investigation demonstrated only subtle cognitive deficits and very few patients were considered impaired on individual tests. There are several possible reasons for this. Firstly, the patients included in this study were at an early stage of their disease when cognitive reserve, functional and structural plasticity, and brain reorganisation may limit cognitive symptoms (Audoin *et al*., 2007b). Other studies have shown cognitive function to be well preserved in groups of RRMS patients with longer disease durations than those subjects investigated in this study (Olivares *et al*., 2005; Schwid *et al*., 2007). It has also been suggested that MS patients

may perform similarly to control subjects in tests which assess the accuracy of performance rather than the time taken to complete the task (Achiron *et al*., 2007; Lazeron *et al*., 2006). Most of the test scores in this study (with the exception of the PASAT, SDMT and Hayling tests) were not time-limited and therefore may not have identified subtle changes. In addition, there may have been a bias in the patients who returned for cognitive testing at five years. It is likely that subjects who were more impaired would be less likely to return or agree to cognitive testing. Four patients in whom baseline lesion volumes and first year brain atrophy rate could be quantified did not have cognitive assessment at five year follow-up and were therefore not included in the study. Of these patients, three were unable to come in for testing due to disability and it is likely that these patients had worse cognition. Although no significant difference in disability was found between those subjects that returned for cognitive testing and those that did not, this analysis may have been limited by the small numbers of patients who did not return for assessment giving the analysis limited power.

Despite the minimal impairment in most subjects (which reduced the power to determine an association), this study found that brain atrophy rate was the only independent predictor of overall cognitive test score at five year follow-up, when taking into account clinical and demographic variables and, importantly, T1 and T2 lesion load. This finding supports the idea that whole brain atrophy is an integral measure of diffuse damage including cortical neuronal loss. It also supports the hypothesis that brain atrophy is a relevant marker of the disease that may aid in monitoring the progression of disease and predicting future disability. Moreover, and perhaps surprisingly, T2 lesion volume was not predictive of any of the composite cognitive scores. This implies that diffuse neuroaxonal damage plays a more important role in future disability in MS than T2 lesion volumes. Arguably lesion volumes are measures of focal transient damage which do not take into account longer term global pathology.

Interestingly, T1 lesion volume, not brain atrophy rate, was found to be predictive of deficits in attention and information processing speed. Patients were most commonly impaired on tests in this domain, in agreement with previous studies showing deficits in this area early in the disease (Achiron & Barak, 2003; Feuillet *et al*., 2007; Olivares *et al*., 2005). Long axonal fibres connecting cerebral regions are contained within the WM and it is thought that demyelination leads to slowing in the speed of neuronal conduction. This could explain the predictive value of T1 lesions which are considered to represent chronic

demyelination and tissue damage. It may be that WM atrophy rather than T1 lesion volume or whole brain atrophy could be a better predictor in this domain, as diffuse pathological changes are taken into account in an area more relevant to deficits in processing speed. A cross-sectional study has previously found that WM volume was the best predictor of mental processing speed and working memory (over GM and lesion volumes) in a group of 40 MS patients (Sanfilipo *et al*., 2006). Area of the CC has also been associated with deficits in information processing speed (Lin *et al*., 2007; Pelletier *et al*., 2001; Rao *et al*., 1989). Furthermore, studies have found ventricular size and enlargement to be associated with performance in the PASAT and SDMT (Benedict *et al*., 2002; Benedict *et al*., 2006; Christodoulou *et al*., 2003; Jasperse *et al*., 2007c), and it has been hypothesised that this ventricular enlargement is due to damage of periventricular WM tracts. In one study, the BPF was only significantly associated with cognitive performance when third ventricular width was excluded from the analysis (Benedict *et al*., 2004).

Only one patient was considered impaired on tests of memory but it has been suggested that memory may be one of the earliest cognitive domains to be affected in MS (Amato *et al*., 2007; Christodoulou *et al*., 2003; Deloire *et al*., 2005; Feuillet *et al*., 2007; Piras *et al*., 2003). Performance on the memory tests was significantly predicted by brain atrophy rate in the first year from baseline. Functions mediated by widely dispersed cortical regions, such as memory, are likely to be related to global cortical damage and disruption of connections between cortical associative areas and cortical/subcortical structures. Brain atrophy is a global measure and so subjects with widespread atrophy are more likely to have cortical dysfunction in multiple areas. MRI studies have shown GM and cortical volumes to be correlated significantly with verbal memory (Amato *et al*., 2004; Benedict *et al*., 2006; Portaccio *et al*., 2006; Sanfilipo *et al*., 2006). At a more focal level, lesions have recently been shown to occur frequently within the hippocampus, a structure thought to play an important role in memory function (Geurts *et al*., 2007; Roosendaal *et al*., 2008).

Over a quarter of patients showed significant IQ decline in at least one IQ measure, but this was not predicted by any MR measure. However premorbid IQ was shown to be related to scores in tests of attention and information processing speed, and executive function which may be evidence of cognitive reserve as has been observed in previous studies (Corral *et al*., 2006). Similarly to IQ decline, performance on tests of executive

function was not predicted by any MR measure. However, high anxiety ratings were associated with poorer executive function. Increased anxiety may use available cognitive capacity in this domain, thus lessening capacity on more demanding tasks. It may also be that impairments in these areas generally take longer to manifest and that impairment is related to many years of pathological damage.

Despite the significant predictive value of brain atrophy rate and T1 lesion volume in certain cognitive domains, they explained only a small proportion of the variance in test scores at five year follow-up. This may be due to several factors. Firstly, as already mentioned, cognitive reserve and structural and functional reorganisation may alter the relationship between brain damage and cognitive performance. Secondly, it is likely that the location of pathology will affect the cognitive changes observed. Therefore, regional measures of atrophy may be more sensitive in predicting cognitive deficits in certain domains. Cross-sectional studies found that measures of temporal lobe atrophy accounted for more variance in tests assessing verbal and spatial memory dysfunction than whole brain atrophy (Benedict *et al*., 2005), and that superior frontal lobe atrophy predicted impairment in verbal learning, spatial learning, attention and conceptual reasoning (Benedict *et al*., 2002; Locatelli *et al*., 2004). These results were corroborated recently in the first study to investigate associations between regional GM volume and neuropsychological function (Tekok-Kilic *et al*., 2007). It found that left frontal GM atrophy was associated with tests of verbal memory whilst right frontal GM atrophy was associated with impairment in visual and working memory.

A limitation of the current study is the small sample size and the results of this study need to be confirmed in larger cohorts of patients. This will also allow the inclusion of other clinical and MRI criteria which may be important predictors of future cognitive status, such as relapse rate or measures from other MR modalities. In addition, these patients did not have baseline cognitive assessments, which would have allowed a more precise assessment of cognitive deterioration and association of the development of cognitive impairment with early MRI markers of disease.

In summary, these results suggest that early brain atrophy rate may be predictive of future cognitive deficits in patients with RRMS. It appears to be a more relevant indicator of prognosis than MRI measures of T2 lesions, as cognitive performance requires the integrity of functional networks between different brain areas. However, the individual

predictive value of GM, WM and regional atrophy rates for future cognitive impairment need to be determined. This could allow identification of patients early in the disease course who are most likely to develop cognitive problems that may impact on daily living, and as such allow preventative and support strategies to be established prior to significant impairment.

9.5 Chapter conclusions

This chapter has investigated the relationship of brain atrophy early in the course of disease to the future development of clinical disability and cognitive impairment in CIS and RRMS. Mean brain atrophy rate is significantly increased from clinical presentation in subjects who are later diagnosed with MS, whilst in those who remain with a clinically isolated syndrome it is close to zero. This finding suggests that brain atrophy may be occurring prior to clinical presentation and that brain atrophy could be used as a marker of disease progression from the very earliest stages of disease. Whilst this study confirmed that the presence of T2 lesions at initial clinical presentation is highly predictive of whether a subject will be diagnosed subsequently with MS, brain atrophy rate during the first year after presentation was found to be independently predictive of the risk of a diagnosis of MS at five year follow-up. This highlights the relevance of brain atrophy to clinical progression and prognosis, but it is premature to suggest that brain atrophy could be used for diagnostic purposes in individual patients.

This chapter has also shown that brain atrophy rate early in the course of relapsing remitting disease is predictive of disability rated by the MSFC, and cognitive impairment. This is further evidence that neuroaxonal loss early in the disease course is of clinical relevance. As such, these results confirm the importance of brain atrophy measurements as a marker of disease progression. However longer follow-up studies are required that include more patients, to explore in more detail the relationship of brain atrophy and other MR markers of disease to clinical disability and cognition. These studies may help to elucidate the mechanisms relating pathology, MR markers and clinical disability in MS.

10 Thesis conclusions

This thesis investigated the application of a wide range of image analysis techniques for measurement of brain atrophy, and their relative potential to monitor disease progression and assess therapeutic efficacy in MS. Despite increasing application of brain atrophy as an outcome measure in trials of putative disease-modifying treatments for MS, the most appropriate technique for this purpose has not been established. It may be that different techniques have value in different cohorts or with different acquisitions and clinical questions. However given that there are a number of potential measures it would be worth optimising and comparing techniques for possible use in common clinical trial scenarios. Furthermore, robust sample size calculations for using brain atrophy rate as an outcome measure have not been performed. The techniques investigated in this thesis range in their methodology and automation, and include those that have previously been applied widely to MS subjects and those that are novel in MS studies.

The focus of this thesis was therefore to assess the most powerful measures of brain atrophy in CIS and RRMS patients who were early in the course of their disease and typical of the subjects who would be included in future treatment trials. Robust sample size calculations were performed to determine the number of RRMS subjects that would be required for a placebo-controlled trial of a potential disease-modifying treatment. Lastly, this thesis evaluated the prognostic value of brain atrophy rate early in the disease for future clinical disease progression.

10.1 Assessment of brain atrophy measurement techniques

In order to assess the relative advantages and disadvantages of atrophy measurement techniques a range of factors were considered including the sensitivity, precision, robustness, reproducibility and degree of operator input required. Measurement error must be minimised as it can reduce sensitivity to change, particularly when atrophy rates are low, as they may be in the early stages of disease, or treatment effects small. With this in mind, automated techniques that directly quantify brain atrophy following registration are attractive. These techniques avoid quantification of brain volumes at serial time-points when errors associated with obtaining each volume may be additive. However such automated techniques often consist of a series of processes and the output of each need to be examined carefully to ensure the general accuracy and precision of the method. In addition, these techniques may not allow for differences in the appearance of images (e.g.

contrast), and selection of the optimal acquisition and processing parameters for a specific dataset may increase the validity of measurements. The initial part of this thesis focussed on the assessment and improvement of brain atrophy measures using two such automated registration-based techniques, the BBSI and SIENA, in MS subjects.

Small, but significant amounts of intensity inhomogeneity between serial images acquired using an identical acquisition protocol on the same scanner may occur during longitudinal studies. Differential bias correction between serial images appeared to increase intensity homogeneity between them, and can significantly increase measurement precision by the BBSI. Application of intensity inhomogeneity correction on single images was also described. Although the BBSI appeared to be robust to small amounts of inhomogeneity within individual images, correction of this artefact may become more relevant as an increasing number of MRI studies are performed on 3T scanners and inhomogeneity may be more prominent. Improvement of acquisitions prior to BBSI atrophy quantification, using fully automated processes that require minimal operator time, such as DBC, can provide significant increases in statistical power, suggesting that these processing steps should be performed routinely.

Application of the BBSI to two FSPGR sequences was described. A small reduction in the variance of measures was observed using the acquisition with marginally larger voxels. However this was non-significant and it was concluded that the BBSI was robust to small differences in acquisition protocol. Interestingly, averaging BBSI measurements obtained from the two FSPGR sequences increased (albeit non-significantly) measurement precision. Averaging two volumetric images has been used in this thesis in an attempt to increase the SNR and provide an image that may allow more sensitive and precise atrophy measurement by the BBSI. This work showed that despite increasing SNR significantly, brain atrophy measurements were not improved and it was concluded that to benefit atrophy measurements significantly, images obtained using exactly the same acquisition parameters, are required. In addition it may be that greater than two images are necessary, and this would be impractical in the context of large clinical trials. Therefore, there was no strong evidence to suggest a benefit from combining two scans at each MRI session for atrophy measurement in a given patient.

Brain atrophy rates quantified by the BBSI can vary considerably depending on the processing parameters used for analysis. Altering the intensity window over which the

BBSI is calculated, can increase brain atrophy measurements. Although measurement precision is reduced in association with this increase, the changes are not proportional and marginal gains in statistical power can be obtained. However the best method by which to assess the optimal intensity window needs to be determined.

SIENA is often applied to "two-dimensional" images with limited resolution in one plane. This thesis showed that high resolution three-dimensional images could be used for reliable brain atrophy quantification and should be considered for future brain atrophy studies in MS. Accurate automated brain extraction can be achieved on these images when processing parameters are carefully selected, whilst manual editing was required on images with lower resolution. The use of affine registration of a template brain mask to high-resolution images, prior to atrophy quantification by SIENA, was shown to produce results consistent with those obtained following generation of subject-specific brain masks. SIENA appears to be robust to small errors in the initial brain extraction. This technique has potential for future analysis of brain atrophy in clinical trials, as generation of brain masks does not rely on image contrast or subject positioning within the FOV. Validation of the technique is required based on different volumetric acquisitions to determine the potential application of this method. Further work to assess ways in which the accuracy of a template brain mask could be improved should be performed, for example non-linear registration of the template to target images, or use of a template library.

High resolution volumetric images may not only provide an easier and more automated means of initial brain extraction than "2D" images, but a systematic bias in atrophy quantification may occur on these two acquisitions. Brain atrophy may be underestimated by SIENA on 2D acquisitions. However higher measurement variability using volumetric acquisitions resulted in the two sequences having similar statistical power. Further investigation of these findings in patients with greater atrophy rates, and application to 2D and 3D images acquired using different parameters or from different scanners needs to be performed.

Direct comparison of brain atrophy rates quantified using different techniques in the same control and MS subjects highlighted some substantial differences in measurements. In addition to being automated, it was determined that the "direct" (registration-based) measures of whole brain atrophy, SIENA and the BBSI, gave significantly better

precision of atrophy measurements than methods based on the segmentation and subtraction of volumes at serial time-points. Atrophy measurements from the BBSI and SIENA correlated well, although SIENA had marginally more statistical power than the BBSI. Another automated technique, SPM segmentation of brain tissue fractions, showed BPF and GMF to have the highest percentage volume changes of all methods. Unlike the other methods investigated, this method provides an indication of pathology in both GM and WM, and as such is useful in increasing understanding of disease mechanisms. However the technique was affected by high measurement variability and was not as robust or reliable as the other techniques, and had much lower statistical power to track atrophy. Automated segmentation techniques such as SPM may be limited by the accuracy of tissue segmentation and the appearance of images is more likely to affect the performance of the algorithm than semi-automated segmentation methods that allow manual adjustment of regions. Optimisation of acquisitions for this technique could improve the results and further work is required to assess the value of other regional atrophy measures that may be applicable for monitoring disease progression in MS.

Measures of ventricular enlargement were an exception to the generally inferior performance of segmentation and subtraction methods. Compared with whole brain measures, the simple (no complex folding of gyri) high contrast boundary between brain and ventricular CSF, and use of a standard intensity threshold for segmentation, mean that there are fewer subjective decisions about the boundaries, and therefore segmentation is highly reproducible and precise. Application of a previously untried method in MS, the VBSI, was described to register local ventricular regions and automatically quantify ventricular enlargement. It was found that there was no significant improvement in the precision of measurement as assessed by comparison with semi-automated segmentation of regions, and underestimation of ventricular enlargement at higher values was apparent. Scan-rescan analysis showed it to be less reliable than semi-automated segmentation and subtraction of volumes, and with similar effect sizes obtained by segmentation and the VBSI, there was no evidence to suggest that the VBSI should be used instead of calculation of segmented ventricle volume difference.

Measures of CCV, which attempt to capture the change occurring in and around the lateral ventricles, appear limited due to application of this technique to “2D” CSE acquisitions. Differences in slice selection and positioning, and possible problems with

interpolation of low resolution images following registration, lead to poor measurement precision and robustness.

To aid reliable quantification of brain atrophy rates, images should ideally be consistent over time. Changes to scanner hardware and software can lead to changes in images that are incorrectly interpreted as brain volume change by some segmentation techniques that are driven by tissue contrast. Altering methodology to correct for these changes in images is one possible solution, but may not adequately compensate for scanner-related changes, and it is impossible to know the true value that should be obtained. Regular scanning of control subjects allows changes over time, and those related to specific changes in MRI hardware and software, to be determined. Regression analysis provides a means by which to estimate and correct for upgrade-related changes in volume. Future work could assess registration based methods and histogram correction of scans combined with atrophy measurement techniques that are driven less by absolute tissue contrast and may be more robust to such changes in acquisitions.

10.2 Sample size calculations

Sample size calculations for RRMS patients were performed for the techniques that had been identified as the more effective techniques for measurement of brain atrophy, namely SIENA, the BBSI and ventricular enlargement. Using a sensitive and precise measurement technique, brain atrophy rate could be a practical addition to outcome measures in clinical trials for MS. The number of RRMS subjects required in each treatment arm for a placebo-controlled trial of a potential disease-modifying treatment was estimated to be comparable to or less than that required for studies using clinical outcome measures, currently the gold standard for phase III trials. SIENA was the most powerful measure, with estimates that only 123 subjects were required per treatment arm to show a 30% slowing in atrophy rate over two years. Over a two year trial no significant difference in the number of subjects required was observed between SIENA, the BBSI and ventricular enlargement, but comparison with a less precise measure of brain atrophy (segmented brain volume difference) highlighted the importance of using a less variable measurement technique. Reduced variability minimises the number of subjects that would need to be exposed to treatments that may be ineffective and have side effects, and reduce the cost of clinical trials. The cost of a trial could potentially also be reduced by shorter follow-up intervals. However it was shown that sample sizes were significantly increased with short trial duration. Furthermore, the mechanism of treatment needs to be considered

to make sure that a delayed effect on disease progression is not overlooked due to a short trial duration.

10.3 Clinical findings

Throughout this thesis application of atrophy measurement techniques to control subjects, subjects presenting with a CIS and patients with RRMS have been described. Brain atrophy is not specific for pathology and represents a net (integral) effect. Transient inflammation, remyelination and gliosis, amongst other factors, may have a confounding effect on brain atrophy measurements and future studies need to investigate the relationship between brain atrophy and these other pathological features of MS. Despite this limitation, a comparison of atrophy rates in early RRMS subjects showed significant differences in rates of whole brain and GM atrophy, and expansion of the lateral ventricles. When global brain atrophy was studied over three years it was concluded that there was no consistent acceleration or deceleration in rate, which might be expected in an inflammatory disease particularly if there is a confound of different treatment being taken during the study, but longer MRI follow-up is required in order to confirm this finding.

This study showed clearly that brain atrophy rates are increased from first clinical presentation. Evidence of significantly increased rates of atrophy in CIS subjects who go on to develop MS relative to control subjects and those who remain stable was presented. Subjects with higher brain atrophy rates from first presentation are at a significantly increased risk of development of MS, and this measure may therefore aid in the prognosis of patients presenting with a CIS. The prognostic value of brain atrophy in CIS patients is independent of that gained from the detection of T2 lesions on MRI at presentation. Further investigation of the value of brain atrophy rates in CIS patients needs to be investigated in larger cohorts.

Evidence from the studies presented in Chapter 9 suggested that rates of atrophy in early RRMS were predictive of future disability and, importantly, of deficits in cognition, independent of the predictive value of MRI lesion measures. Brain atrophy rate was a significant independent predictor of disability five years later, as measured by the MSFC, a measure of ambulation, arm/hand function and cognition. Subtle cognitive impairment was observed in RRMS patients when cognitive testing was performed in subjects at five year follow-up. Brain atrophy rate was found to be the most significant independent predictor of overall cognitive score and memory function at follow-up. Future studies

investigating regional brain atrophy and cognitive deficits may reveal specific associations that could more reliably predict future cognitive impairment.

10.4 Summary

Brain atrophy is a clinically relevant marker of disease progression in MS. This thesis investigates a range of brain atrophy measurement techniques in MS. It shows that techniques vary considerably in their sensitivity, precision, robustness and reproducibility. Choosing appropriate atrophy measurement techniques, it is possible to reliably detect treatment effects using brain atrophy as an outcome measure. The possibility of improving the reliability of these measures and thereby contributing to improved drug discovery in MS is an important and feasible aim. Despite a number of important caveats and cautions it appears likely that treatments that slow atrophy progression, especially over longer studies, may truly be reducing the long term burden of disability in this devastating disease.

Appendices

Appendix 1: Kurtzke Expanded Disability Scale

The Kurtzke Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) is a method of quantifying disability in MS. The EDSS quantifies disability in eight functional systems (FS) and allows neurologists to assign a functional system score in each of these.

The function systems are:

- Pyramidal
- Cerebellar
- Brainstem
- Sensory
- Bowel and bladder
- Visual
- Cerebral
- Other

EDSS steps 1.0 to 4.5 refer to people with MS who are fully ambulatory, and the precise step is defined by the functional system scores. EDSS steps 5.0 to 9.5 are defined by the impairment to ambulation.

Score	*Description*
0	Normal neurological examination (all grade 0 in FS).
1.0	No disability, minimal signs in one FS.
1.5	No disability, minimal signs in more than one FS.
2.0	Minimal disability in one FS.
2.5	Mild disability in one FS or minimal disability in two FS.
3.0	Moderate disability in one FS, or mild disability in three or four FS. Fully ambulatory.
3.5	Fully ambulatory but with moderate disability in one FS and more than minimal disability in several others.
4.0	Fully ambulatory without aid, self-sufficient, up and about some 12 hours a day despite relatively severe disability; able to walk without aid or rest some 500 metres.
4.5	Fully ambulatory without aid, up and about much of the day, able to work a full day, may otherwise have some limitation of full activity or require minimal assistance; characterised by relatively severe disability; able to walk without aid or rest some 300m.
5.0	Ambulatory without aid or rest for about 200 metres; disability severe enough to impair full daily activities (work a full day without special provisions).
5.5	Ambulatory without aid or rest for about 100 metres; disability severe enough to preclude full daily activities.
6.0	Intermittent or unilateral constant assistance (cane, crutch, brace) required to walk about 100 metres with or without resting.
6.5	Constant bilateral assistance (canes, crutches, braces) required to walk about 20 metres without resting.
7.0	Unable to walk beyond approximately five metres even with aid, essentially restricted to wheelchair; wheels self in standard wheelchair and transfers alone; up and about in wheelchair some 12 hours a day.
7.5	Unable to take more than a few steps; restricted to wheelchair; may need aid in transfer; wheels self but cannot carry on in standard wheelchair a full day; may require motorised wheelchair.
8.0	Essentially restricted to bed or chair or perambulated in wheelchair, but may be out of bed itself much of the day; retains many self-care functions; generally has effective use of arms.
8.5	Essentially restricted to bed much of day; has some effective use of arms; retains some self care functions.
9.0	Confined to bed; can still communicate and eat.
9.5	Totally helpless bed patient; unable to communicate effectively or eat/swallow.
10.0	Death due to MS.

Appendix 2: The Multiple Sclerosis Functional Composite

The Multiple Sclerosis Functional Composite (MSFC) (Fischer *et al*., 1999) is a multi-dimensional measure which reflects the varied clinical expression, including cognitive function, of MS across patients and over time, and allows for the dimensions to change relatively independently over time. The MSFC is comprised of:

1. Timed 25-foot walk (TW)
2. Nine-hole peg test (9HPT)
3. Paced auditory serial addition test (PASAT-3 version).

The results from each of these three tests are transformed into z-scores and averaged to yield a composite score for each patient at each time-point.

Timed 25-foot walk

The TW is a quantitative measure of lower extremity function. The patient is directed to one end of a clearly marked 25-foot course and is instructed to walk 25 feet as quickly as possible, but safely. The task is immediately administered again by having the patient walk back the same distance. Patients may use assistive devices when doing this task. The time taken for each trial is recorded.

Nine-hole peg test

The 9HPT is a quantitative measure of upper extremity (arm and hand) function. Both the dominant and non-dominant hands are tested twice. The patient must pick up pegs one at a time, using one hand only, and put them into holes as quickly as possible until all the holes are filled. Then without pausing, the patient must remove the pegs one at a time and return them to the container as quickly as possible. The time taken for each trial is recorded.

Paced auditory serial addition test

The PASAT is a measure of cognitive function that specifically assesses auditory information processing speed and flexibility, as well as calculation ability. It requires the patient to listen to a series of single digit numbers on a CD which are presented at a rate of one every three seconds. The patient is required to listen to the first two numbers, add them up and tell the tester the answer. When the next number is presented the patient must

add it to the number preceding it. The patient continues to add each number presented to the one preceding it. The number of correct responses is recorded.

MSFC score

TW: the times from the two trials are averaged.

Nine-HPT: the two trials for each hand are averaged, converted to reciprocals of the mean times for each hand, and then the two reciprocals are averaged.

PASAT: the number correct from the PASAT-3.

Z-scores are created for each of the component scores, using test results from the baseline visit from all patients in a study cohort. Thus the z-score is a standardised number representing how close a test result is to the mean of a standard reference population to which the result is compared. Overall MSFC z-score is calculated as follows:

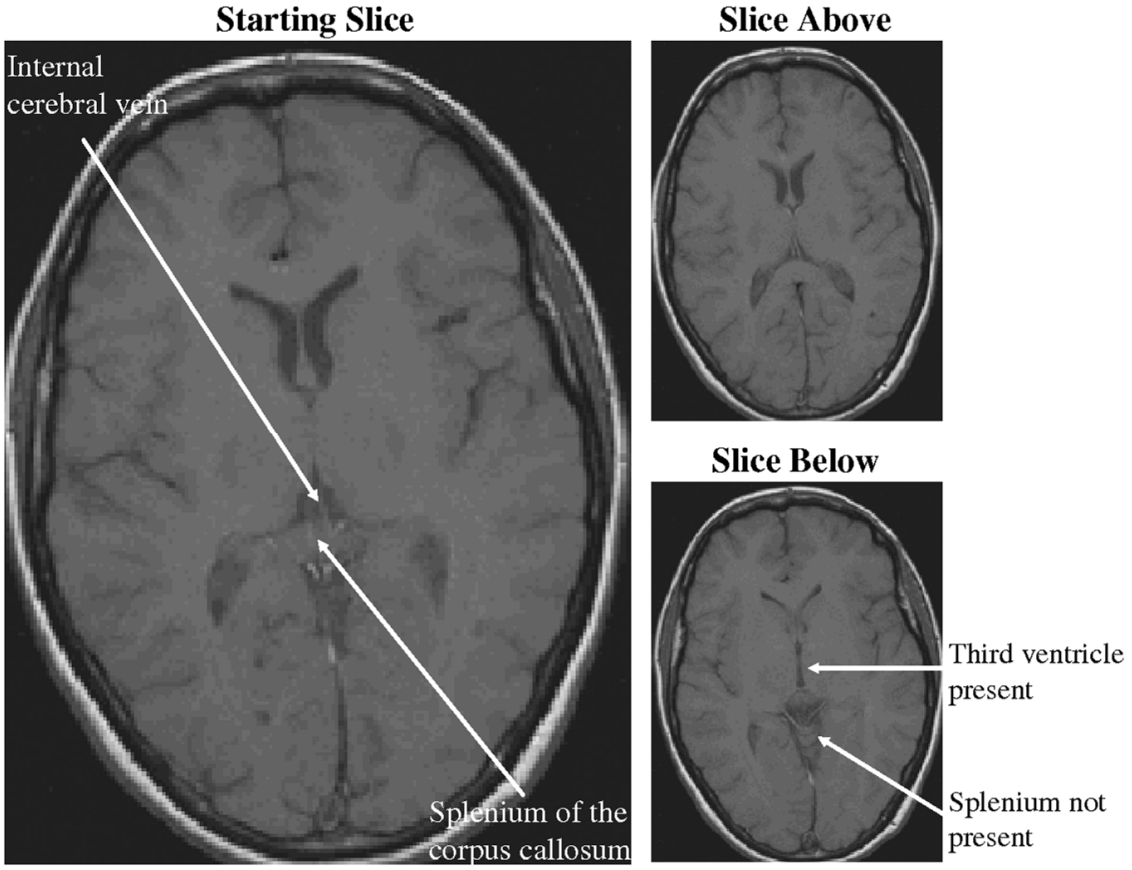
MSFC z-score =

[{(average(1/9HPT) – baseline mean (1/9HPT))/baseline SD(1/9HPT)}

+ {-(average TW – baseline mean TW)/baseline SD TW}

+ {(PASAT-3 – baseline mean PASAT-3)/baseline SD PASAT-3}] /3

Appendix 3: Starting slice for Losseff technique



Proceeding superiorly, the slice where the third ventricle had disappeared and the splenium of the corpus callosum had appeared, was determined. In addition, the presence of the internal cerebral veins were be used as a guide to the best slice to start measurements especially on images acquired with gadolinium where they were particularly visible.

List of abbreviations

9HPT	Nine-hole peg test
AIR	Automated image registration
ApoE	Apolipoprotein E
BBSI	Brain boundary shift integral
BET	Brain extraction tool
BICCR	Brain to intracranial capacity ratio
BPF	Brain parenchymal fraction
BVD	Brain volume difference
CC	Corpus callosum
CCV	Central cerebral volume
CI	Confidence interval
CIS	Clinically isolated syndrome
CNR	Contrast-to-noise ratio
CNS	Central nervous system
Cr	Creatine
CSE	Conventional spin echo
CSF	Cerebrospinal fluid
CV	Coefficient of variation
DBC	Differential bias correction
DTI	Diffusion tensor imaging
DW-MRI	Diffusion weighted magnetic resonance imaging
EDSS	Expanded disability status scale
FAST	FMRIB's automated segmentation tool
FCP	Fuzzy connected principles
FLAIR	Fluid attenuated inversion recovery
FLIRT	FMRIB's linear image registration tool
FMRIB	Functional magnetic resonance imaging of the brain
FOV	Field of view
FSE	Fast spin echo
FSL	FMRIB's software library
FSPGR	Fast spoiled gradient recall
FS	Functional system
Gd-DTPA	Gadolinium diethyltriaminepentaacetic acid

GM	Grey matter
GMF	Grey matter fraction
GRASS	Gradient recalled at steady state
^{1}H-MRS	Proton magnetic resonance spectroscopy
HADS	Hospital anxiety and depression scale
ICC	Intraclass correlation coefficient
IQR	Interquartile range
LR	Likelihood ratio
MIDAS	Medical image display and analysis system
MNI	Montreal Neurological Institute
MPRAGE	Magnetisation-prepared rapid acquisition gradient echo
MRI	Magnetic resonance imaging
MS	Multiple sclerosis
MSFC	Multiple sclerosis functional composite
MT-MRI	Magnetisation transfer-magnetic resonance imaging
MTR	Magnetisation transfer ratio
N3	Nonparametric nonuniform intensity normalisation
NAA	N-acetylaspartate
NAGM	Normal appearing grey matter
NART	National adult reading test
NAWM	Normal appearing white matter
NBV	Normalised brain volume
OR	Odds ratio
PASAT	Paced Auditory Serial Addition Test
PBVC	Percentage brain volume change
PD	Proton density
PPMS	Primary progressive multiple sclerosis
ROI	Region of interest
RRMS	Relapsing remitting multiple sclerosis
SDMT	Symbol digits modalities test
SIENA	Structural image evaluation, using normalisation, of atrophy
SIENAX	Structural image evaluation, using normalisation, of atrophy–cross-sectional
SNR	Signal-to-noise ratio
SPGR	Spoiled gradient recall
SPM	Statistical parametric mapping

SPMS	Secondary progressive multiple sclerosis
SWM	Spatial working memory
TE	Time to echo
TI	Time to inversion
TR	Time to repeat
TW	Twenty-five foot timed walk
VBM	Voxel based morphometry
VBSI	Ventricular boundary shift integral
VE	Ventricular enlargement
VVD	Ventricular volume difference
WAIS	Weschler adult intelligence scale
WBV	Whole brain volume
WM	White matter
WMF	White matter fraction

Publications

Published peer-reviewed papers based on results described in this thesis are listed. I am grateful to the individuals involved in these publications, whose contributions are detailed below.

Anderson VM, Fox NC, Miller DH, 2006.Magnetic resonance imaging measures of brain atrophy in multiple sclerosis. *Journal of Magnetic Resonance Imaging*,23, p.605-618. (Chapters 1 and 2).
I was responsible for drafting the review, which was revised by Nick Fox and David Miller.

Anderson VM, Fernando KTM, Davies GR, Rashid W, Frost C, Fox NC, Miller DH, 2007. Cerebral atrophy in patients presenting with clinically isolated syndromes and early relapsing-remitting multiple sclerosis: a comparison of registration-based methods. *Journal of Neuroimaging*, 17, p.61-68. (Chapter 6.2).
Patients with RRMS and control subjects were recruited and assessed by Gerard Davies and Waqar Rashid. Patients with CIS were recruited and assessed by Kryshani Fernando. I carried out brain segmentations, BBSI and SIENA analysis on MRI. Statistical advice was provided by Chris Frost. I was responsible for analysis and drafting of the paper, which was revised by Chris Frost, Nick Fox and David Miller.

Anderson VM, Bartlett JW, Fox NC, Fisniku L, Miller DH, 2007. Detecting treatment effects on brain atrophy in relapsing remitting multiple sclerosis: sample size estimations. *Journal of Neurology*, 254, p. 1588-1594. (Chapter 7).
Patients with RRMS and control subjects were assessed by Leonora Fisniku. Jonathan Bartlett provided statistical advice and performed data modelling and sample size estimations. I performed brain segmentations, BBSI and SIENA analyses and output data for statistical analyses. I was responsible for drafting the paper, which was revised by Jonathan Bartlett, Nick Fox and David Miller.

Summers M, Fisniku L, **Anderson VM**, Miller DH, Cipolotti L, Ron MA, 2007. Predicting cognitive impairment in relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, 13, p. 1-8. (Chapter 9.4).
Mary Summers was responsible for neuropsychological assessment of patients, statistical analysis and drafting of the paper. Clinical assessments were performed by Leonora Fisniku. I was responsible for brain atrophy measurements on these patients and contributed to revisions of the paper. David Miller, Lisa Cipolotti and Maria Ron were responsible for supervising the project.

Acknowledgements

There are many people who made this work possible. Firstly I am grateful to all the patients and control subjects who gave their time for the research studies included in this thesis. I would like to thank the radiographers in the MS NMR Unit who acquired the scans used for this work and advised me on MRI: Ros Gordon, Chris Benton, Kelly Wimpey and Dave McManus. I am extremely grateful to the clinical fellows within the MS NMR Research Unit who spent many hours assessing subjects, supervising MRI sessions and providing me with clinical data, particularly Declan Chard, Gerard Davies, Kryshani Fernando, Leonora Fisniku, Waqar Rashid and Jo Swanton. In addition, the lesion volume data used in Chapter 9 was generated by Kryshani (CIS patients) and Waqar (RRMS patients), and I would like to thank them both for sharing this data.

I would also like to thank Mary Summers who performed neuropsychological assessments on patients and contributed the data for Chapter 9.4. She was also responsible for the descriptions of the neuropsychological tests included in this chapter. Sophie Penny has also been a great help, answering my queries on this data following Mary's departure.

I am extremely grateful for the statistical advice given by Chris Frost, Jonathan Bartlett and Dan Altmann. Jonathan performed the data modelling and helped with sample size calculations for Chapter 7 and Dan helped to develop the statistical model presented in Chapter 8.2.

Thanks must also go to other people I have worked with at the MS NMR Research Unit and Dementia Research Centre who have directly helped with the work that formed this thesis: Caroline Sherrard, Tracey Pepple, Jo Foster, Nicola Hobbs, Tom Yeatman, and Richard Boyes. Also, the people I have worked alongside, helped me indirectly, and advised and supported me throughout my PhD: Jo Barnes, Susie Henley, Shona Price, Ali Bendriss, Abdel Douiri, Gordon Ingle, Jon Jackson, Richard Lanyon, Musib Siddique, Derek Soon, Jon Steele and Dan Tozer. I would also like to thank the administrative support, particularly Gillian Barley who helped to arrange supervisions, and Moira Young, Anne Parnell and Lynn Maslen.

I am extremely grateful to my supervisors Professor Nick Fox and Professor David Miller, and to Dr Rachael Scahill, who have given me expert guidance and support from beginning to end, and made this PhD possible.

I would also like to thank the Multiple Sclerosis Society of Great Britain and Northern Ireland who provided funding for this work.

Lastly, I'd like to thank Rich for his patience and support which has helped me so much throughout this thesis.

References

Achiron, A. & Barak, Y., 2003. Cognitive impairment in probable multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 74, p. 443-446.

Achiron, A., Doniger, G.M., Harel, Y., Appleboim-Gavish, N., Lavie, M. & Simon, E.S., 2007. Prolonged response times characterize cognitive performance in multiple sclerosis. *European Journal of Neurology*, 14, p.1102-1108.

Achiron, A., Gicquel, S., Miron, S. & Faibel, M., 2002. Brain MRI lesion load quantification in multiple sclerosis: a comparison between automated multispectral and semi-automated thresholding computer-assisted techniques. *Magnetic Resonance Imaging*, 20, p. 713-720.

Adalsteinsson, E., Langer-Gould, A., Homer, R.J., Rao, A., Sullivan, E.V., Lima, C.A., Pfefferbaum, A. & Atlas, S.W., 2003. Gray matter N-acetyl aspartate deficits in secondary progressive but not relapsing-remitting multiple sclerosis. *AJNR American Journal of Neuroradiology*, 24, p. 1941-1945.

Agosta, F., Rovaris, M., Pagani, E., Sormani, M.P., Comi, G. & Filippi, M., 2006. Magnetization transfer MRI metrics predict the accumulation of disability 8 years later in patients with multiple sclerosis. *Brain*, 129, p. 2620-2627.

Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A. & Moriarty, T., 2002. A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transactions on Medical Imaging*, 21, p. 193-199.

Amato, M.P., Bartolozzi, M.L., Zipoli, V., Portaccio, E., Mortilla, M., Guidi, L., Siracusa, G., Sorbi, S., Federico, A. & De Stefano, N., 2004. Neocortical volume decrease in relapsing-remitting MS patients with mild cognitive impairment. *Neurology*, 63, p. 89-93.

Amato, M.P., Ponziani, G., Siracusa, G. & Sorbi, S., 2001. Cognitive dysfunction in early-onset multiple sclerosis: a reappraisal after 10 years. *Archives of Neurology*, 58, p. 1602-1606.

Amato, M.P., Portaccio, E., Goretti, B., Zipoli, V., Battaglini, M., Bartolozzi, M.L., Stromillo, M.L., Guidi, L., Siracusa, G., Sorbi, S., Federico, A. & De Stefano, N., 2007. Association of neocortical volume changes with cognitive deterioration in relapsing-remitting multiple sclerosis. *Archives of Neurology*, 64, p. 1157-1161.

Arnold, J.B., Liow, J.S., Schaper, K.A., Stern, J.J., Sled, J.G., Shattuck, D.W., Worth, A.J., Cohen, M.S., Leahy, R.M., Mazziotta, J.C. & Rottenberg, D.A., 2001. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *NeuroImage*, 13, p. 931-943.

Ashburner, J. & Friston, K., 1997. Multimodal image coregistration and partitioning - a unified framework. *NeuroImage*, 6, p. 209-217.

Ashburner, J. & Friston, K.J., 2000. Voxel-based morphometry - the methods. *NeuroImage*, 11, p. 805-821.

Audoin, B., Davies, G., Rashid, W., Fisniku, L., Thompson, A.J. & Miller, D.H., 2007a. Voxel-based analysis of grey matter magnetization transfer ratio maps in early relapsing remitting multiple sclerosis. *Multiple Sclerosis*, 13, p. 483-489.

Audoin, B., Davies, G.R., Finisku, L., Chard, D.T., Thompson, A.J. & Miller, D.H., 2006. Localization of grey matter atrophy in early RRMS : A longitudinal study. *Journal of Neurology*, 253, p. 1495-1501.

Audoin, B., Guye, M., Reuter, F., Au Duong, M.V., Confort-Gouny, S., Malikova, I., Soulier, E., Viout, P., Ali-Chérif, A., Cozzone, P.J., Pelletier, J. & Ranjeva, J.P., 2007b. Structure of WM bundles constituting the working memory system in early multiple sclerosis: a quantitative DTI tractography study. *NeuroImage*, 36, p. 1324-1330.

Audoin, B., Ibarrola, D., Malikova, I., Soulier, E., Confort-Gouny, S., Au Duong, M.V., Reuter, F., Viout, P., Ali-Chérif, A., Cozzone, P.J., Pelletier, J. & Ranjeva, J.P., 2007c. Onset and underpinnings of white matter atrophy at the very early stage of multiple sclerosis-a two-year longitudinal MRI/MRSI study of corpus callosum. *Multiple Sclerosis*, 13, p. 41-51.

Audoin, B., Ranjeva, J.P., Au Duong, M.V., Ibarrola, D., Malikova, I., Confort-Gouny, S., Soulier, E., Viout, P., Ali-Chérif, A., Pelletier, J. & Cozzone, P.J., 2004. Voxel-based analysis of MTR images: a method to locate gray matter abnormalities in patients at the earliest stage of multiple sclerosis. *Journal of Magnetic Resonance Imaging*, 20, p. 765-771.

Babbe, H., Roers, A., Waisman, A., Lassmann, H., Goebels, N., Hohlfeld, R., Friese, M., Schröder, R., Deckert, M., Schmidt, S., Ravid, R. & Rajewsky, K., 2000. Clonal expansions of $CD8^+$ T cells dominate the T cell infiltrate in active multiple sclerosis lesions as shown by micromanipulation and single cell polymerase chain reaction. *Journal of Experimental Medicine*, 192, p. 393-404.

Barkhof, F., Filippi, M., Miller, D.H., Scheltens, P., Campi, A., Polman, C.H., Comi, G., Adèr, H.J., Losseff, N. & Valk, J., 1997. Comparison of MRI criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain*, 120, p. 2059-2069.

Barkhof, F., Elton, M., Lindeboom, J., Tas, M.W., Schmidt, W.F., Hommes, O.R., Polman, C.H., Kok, A. & Valk, J., 1998. Functional correlates of callosal atrophy in relapsing-remitting multiple sclerosis patients. A preliminary MRI study. *Journal of Neurology*, 245, p. 153-158.

Barnett, M.H. & Prineas, J.W., 2004. Relapsing and remitting multiple sclerosis: pathology of the newly forming lesion. *Annals of Neurology*, 55, p. 458-468.

Benedict, R.H., Bakshi, R., Simon, J.H., Priore, R., Miller, C. & Munschauer, F., 2002. Frontal cortex atrophy predicts cognitive impairment in multiple sclerosis. *Journal of Neuropsychiatry and Clinical Neurosciences*, 14, p. 44-51.

Benedict, R.H., Bruce, J.M., Dwyer, M.G., Abdelrahman, N., Hussein, S., Weinstock-Guttman, B., Garg, N., Munschauer, F. & Zivadinov, R., 2006. Neocortical atrophy, third ventricular width, and cognitive dysfunction in multiple sclerosis. *Archives of Neurology*, 63, p. 1301-1306.

Benedict, R.H., Weinstock-Guttman, B., Fishman, I., Sharma, J., Tjoa, C.W. & Bakshi, R., 2004. Prediction of neuropsychological impairment in multiple sclerosis: comparison

of conventional magnetic resonance imaging measures of atrophy and lesion burden. *Archives of Neurology*, 61, p. 226-230.

Benedict, R.H., Zivadinov, R., Carone, D.A., Weinstock-Guttman, B., Gaines, J., Maggiore, C., Sharma, J., Tomassi, M.A. & Bakshi, R., 2005. Regional lobar atrophy predicts memory impairment in multiple sclerosis. *AJNR American Journal of Neuroradiology*, 26, p. 1824-1831.

Bermel, R.A., Bakshi, R., Tjoa, C., Puli, S.R. & Jacobs, L., 2002. Bicaudate ratio as a magnetic resonance imaging marker of brain atrophy in multiple sclerosis. *Archives of Neurology*, 59, p. 275-280.

Bermel, R.A., Innus, M.D., Tjoa, C.W. & Bakshi, R., 2003a. Selective caudate atrophy in multiple sclerosis: a 3D MRI parcellation study. *NeuroReport*, 14, p. 335-339.

Bermel, R.A., Sharma, J., Tjoa, C.W., Puli, S.R. & Bakshi, R., 2003b. A semiautomated measure of whole-brain atrophy in multiple sclerosis. *Journal of the Neurological Sciences*, 208, p. 57-65.

Bitsch, A., Schuchardt, J., Bunkowski, S., Kuhlmann, T. & Brück, W., 2000. Acute axonal injury in multiple sclerosis. Correlation with demyelination and inflammation. *Brain*, 123, p. 1174-1183.

Bjartmar, C., Kinkel, R.P., Kidd, G., Rudick, R.A. & Trapp, B.D., 2001. Axonal loss in normal-appearing white matter in a patient with acute MS. *Neurology*, 57, p. 1248-1252.

Bø, L., Geurts, J.J., van der Valk, P., Polman, C. & Barkhof, F., 2007. Lack of correlation between cortical demyelination and white matter pathologic changes in multiple sclerosis. *Archives of Neurology*, 64, p. 76-80.

Bø, L., Vedeler, C.A., Nyland, H., Trapp, B.D. & Mørk, S.J., 2003a. Intracortical multiple sclerosis lesions are not associated with increased lymphocyte infiltration. *Multiple Sclerosis*, 9, p. 323-331.

Bø, L., Vedeler, C.A., Nyland, H.I., Trapp, B.D. & Mørk, S.J., 2003b. Subpial demyelination in the cerebral cortex of multiple sclerosis patients. *Journal of Neuropathology and Experimental Neurology*, 62, p. 723-732.

Bosc, M., Heitz, F., Armspach, J.P., Namer, I., Gounot, D. & Rumbach, L., 2003. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *NeuroImage*, 20, p. 643-656.

Boyes, R.G., Schott, J.M., Frost, C. & Fox, N.C., 2004. Increasing accuracy of atrophy measures from serial MR scans using parameter analysis of the boundary shift integral. *Lecture Notes In Computer Science*, 3217, p. 1036-1037.

Bozzali, M., Cercignani, M., Sormani, M.P., Comi, G. & Filippi, M., 2002. Quantification of brain gray matter damage in different MS phenotypes by use of diffusion tensor MR imaging. *AJNR American Journal of Neuroradiology*, 23, p. 985-988.

Brass, S.D., Narayanan, S., Antel, J.P., Lapierre, Y., Collins, L. & Arnold, D.L., 2004. Axonal damage in multiple sclerosis patients with high versus low expanded disability status scale score. *Canadian Journal of Neurological Sciences*, 31, p. 225-228.

Brex, P.A., Ciccarelli, O., O'Riordan, J.I., Sailer, M., Thompson, A.J. & Miller, D.H., 2002. A longitudinal study of abnormalities on MRI and disability from multiple sclerosis. *New England Journal of Medicine*, 346, p. 158-164.

Brex, P.A., Jenkins, R., Fox, N.C., Crum, W.R., O'Riordan, J.I., Plant, G.T. & Miller, D.H., 2000. Detection of ventricular enlargement in patients at the earliest clinical stage of MS. *Neurology*, 54, p. 1689-1691.

Burgess, P.W. & Shallice, T., 1997. The Hayling and Brixton Tests. Thurston, Suffolk, UK: Thames Valley Test Company.

Cader, S., Johansen-Berg, H., Wylezinska, M., Palace, J., Behrens, T.E., Smith, S. & Matthews, P.M., 2007. Discordant white matter N-acetylasparate and diffusion MRI measures suggest that chronic metabolic dysfunction contributes to axonal pathology in multiple sclerosis. *NeuroImage*, 36, p. 19-27.

Calabrese, M., Atzori, M., Bernardi, V., Morra, A., Romualdi, C., Rinaldi, L., McAuliffe, M.J., Barachino, L., Perini, P., Fischl, B., Battistin, L. & Gallo, P., 2007a. Cortical atrophy is relevant in multiple sclerosis at clinical onset. *Journal of Neurology*, p. 1212-1220.

Calabrese, M., De Stefano, N., Atzori, M., Bernardi, V., Mattisi, I., Barachino, L., Morra, A., Rinaldi, L., Romualdi, C., Perini, P., Battistin, L. & Gallo, P., 2007b. Detection of cortical inflammatory lesions by double inversion recovery magnetic resonance imaging in patients with multiple sclerosis. *Archives of Neurology*, 64, p. 1416-1422.

Calmon, G. & Roberts, N., 2000. Automatic measurement of changes in brain volume on consecutive 3D MR images by segmentation propagation. *Magnetic Resonance Imaging*, 18, p. 439-453.

Camp, S.J., Stevenson, V.L., Thompson, A.J., Miller, D.H., Borras, C., Auriacombe, S., Brochet, B., Falautano, M., Filippi, M., Hérissé-Dulo, L., Montalban, X., Parrcira, E., Polman, C.H., De Sa, J. & Langdon, D.W., 1999. Cognitive function in primary progressive and transitional progressive multiple sclerosis: a controlled study with MRI correlates. *Brain*, 122, p. 1341-1348.

Caon, C., Zvartau-Hind, M., Ching, W., Lisak, R.P., Tselis, A.C. & Khan, O.A., 2003. Intercaudate nucleus ratio as a linear measure of brain atrophy in multiple sclerosis. *Neurology*, 60, p. 323-325.

Caramia, F., Pantano, P., Di Legge, S., Piattella, M.C., Lenzi, D., Paolillo, A., Nucciarelli, W., Lenzi, G.L., Bozzao, L. & Pozzilli, C., 2002. A longitudinal study of MR diffusion changes in normal appearing white matter of patients with early multiple sclerosis. *Magnetic Resonance Imaging*, 20, p. 383-388.

Carone, D.A., Benedict, R.H., Dwyer, M.G., Cookfair, D.L., Srinivasaraghavan, B., Tjoa, C.W. & Zivadinov, R., 2006. Semi-automatic brain region extraction (SABRE) reveals superior cortical and deep gray matter atrophy in MS. *NeuroImage*, 29, p. 505-514.

Cassol, E., Ranjeva, J.P., Ibarrola, D., Mékies, C., Manelfe, C., Clanet, M. & Berry, I., 2004. Diffusion tensor imaging in multiple sclerosis: a tool for monitoring changes in normal-appearing white matter. *Multiple Sclerosis*, 10, p. 188-196.

Cercignani, M., Bozzali, M., Iannucci, G., Comi, G. & Filippi, M., 2001a. Magnetisation transfer ratio and mean diffusivity of normal appearing white and grey matter from patients with multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 70, p. 311-317.

Cercignani, M., Inglese, M., Pagani, E., Comi, G. & Filippi, M., 2001b. Mean diffusivity and fractional anisotropy histograms of patients with multiple sclerosis. *AJNR American Journal of Neuroradiology*, 22, p. 952-958.

Cercignani, M., Symms, M.R., Schmierer, K., Boulby, P.A., Tozer, D.J., Ron, M., Tofts, P.S. & Barker, G.J., 2005. Three-dimensional quantitative magnetisation transfer imaging of the human brain. *NeuroImage*, 27, p. 436-441.

Chard, D.T., Brex, P.A., Ciccarelli, O., Griffin, C.M., Parker, G.J., Dalton, C., Altmann, D.R., Thompson, A.J. & Miller, D.H., 2003. The longitudinal relation between brain lesion load and atrophy in multiple sclerosis: a 14 year follow up study. *Journal of Neurology Neurosurgery and Psychiatry*, 74, p. 1551-1554.

Chard, D.T., Griffin, C.M., McLean, M.A., Kapeller, P., Kapoor, R., Thompson, A.J. & Miller, D.H., 2002a. Brain metabolite changes in cortical grey and normal-appearing white matter in clinically early relapsing-remitting multiple sclerosis. *Brain*, 125, p. 2342-2352.

Chard, D.T., Griffin, C.M., Parker, G.J., Kapoor, R., Thompson, A.J. & Miller, D.H., 2002b. Brain atrophy in clinically early relapsing-remitting multiple sclerosis. *Brain*, 125, p. 327-337.

Chard, D.T., Griffin, C.M., Rashid, W., Davies, G.R., Altmann, D.R., Kapoor, R., Barker, G.J., Thompson, A.J. & Miller, D.H., 2004. Progressive grey matter atrophy in clinically early relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, 10, p. 387-391.

Chard, D.T., Parker, G.J., Griffin, C.M., Thompson, A.J. & Miller, D.H., 2002c. The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology. *Journal of Magnetic Resonance Imaging*, 15, p. 259-267.

Charil, A., Zijdenbos, A.P., Taylor, J., Boelman, C., Worsley, K.J., Evans, A.C. & Dagher, A., 2003. Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis: application to 452 patient data sets. *NeuroImage*, 19, p. 532-544.

Chen, J. & Reutens, D.C., 2005. Inhomogeneity correction for brain magnetic resonance images by rank leveling. *Journal of Computer Assisted Tomography*, 29, p. 668-676.

Chen, J.T., Narayanan, S., Collins, D.L., Smith, S.M., Matthews, P.M. & Arnold, D.L., 2004. Relating neocortical pathology to disability progression in multiple sclerosis using MRI. *NeuroImage*, 23, p. 1168-1175.

Cheng, H. & Huang, F., 2006. Magnetic resonance imaging image intensity correction with extrapolation and adaptive smoothing. *Magnetic Resonance in Medicine*, 55, p. 959-966.

Christodoulou, C., Krupp, L.B., Liang, Z., Huang, W., Melville, P., Roque, C., Scherl, W.F., Morgan, T., MacAllister, W.S., Li, L., Tudorica, L.A., Li, X., Roche, P. & Peyster, R., 2003. Cognitive performance and MR markers of cerebral injury in cognitively impaired MS patients. *Neurology*, 60, p. 1793-1798.

Ciccarelli, O., Brex, P.A., Thompson, A.J. & Miller, D.H., 2002. Disability and lesion load in MS: a reassessment with MS functional composite score and 3D fast FLAIR. *Journal of Neurology*, 249, p. 18-24.

Cifelli, A., Arridge, M., Jezzard, P., Esiri, M.M., Palace, J. & Matthews, P.M., 2002. Thalamic neurodegeneration in multiple sclerosis. *Annals of Neurology*, 52, p. 650-653.

Clark, K.A., Woods, R.P., Rottenberg, D.A., Toga, A.W. & Mazziotta, J.C., 2006. Impact of acquisition protocols and processing streams on tissue segmentation of T1 weighted MR images. *NeuroImage*, 29, p. 185-202.

Cohen, M.S., DuBois, R.M. & Zeineh, M.M., 2000. Rapid and effective correction of RF inhomogeneity for high field magnetic resonance imaging. *Human Brain Mapping*, 10, p. 204-211.

Coles, A.J., Wing, M.G., Molyneux, P., Paolillo, A., Davie, C.M., Hale, G., Miller, D., Waldmann, H. & Compston, A., 1999. Monoclonal antibody treatment exposes three mechanisms underlying the clinical course of multiple sclerosis. *Annals of Neurology*, 46, p. 296-304.

Collins, D., Montagnat, J., Zijdenbos A.P., Evans A.C. & Arnold, D.L., 2001. Automated Estimation of Brain Volume in Multiple Sclerosis with BICCR. *Lecture Notes In Computer Science*, 2082, p. 141-147.

Comi, G., Filippi, M., Barkhof, F., Durelli, L., Edan, G., Fernández, O., Hartung, H.P., Seeldrayers, P., Sørensen, P.S., Rovaris, M., Martinelli, V. & Hommes, O.R., 2001. Effect of early interferon treatment on conversion to definite multiple sclerosis: a randomised study. *Lancet*, 357, p. 1576-1582.

Coombs, B.D., Best, A., Brown, M.S., Miller, D.E., Corboy, J., Baier, M. & Simon, J.H., 2004. Multiple sclerosis pathology in the normal and abnormal appearing white matter of the corpus callosum by diffusion tensor imaging. *Multiple Sclerosis*, 10, p. 392-397.

Corral, M., Rodríguez, M., Amenedo, E., Sánchez, J.L. & Díaz, F., 2006. Cognitive reserve, age, and neuropsychological performance in healthy participants. *Developmental Neuropsychology*, 29, p. 479-491.

Coughlan, A.K. & Hollows, S.E., 1985. The Adult Memory and Information Processing Battery. St James's Hospital, Leeds, UK.

Craner, M.J., Newcombe, J., Black, J.A., Hartle, C., Cuzner, M.L. & Waxman, S.G., 2004. Molecular changes in neurons in multiple sclerosis: altered axonal expression of $Na_v1.2$ and $Na_v1.6$ sodium channels and Na^+/Ca^{2+} exchanger. *Proceedings of the National Academy of Sciences U S A*, 101, p. 8168-8173.

Cutter, G.R., Baier, M.L., Rudick, R.A., Cookfair, D.L., Fischer, J.S., Petkau, J., Syndulko, K., Weinshenker, B.G., Antel, J.P., Confavreux, C., Ellison, G.W., Lublin, F., Miller, A.E., Rao, S.M., Reingold, S., Thompson, A. & Willoughby, E., 1999. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*, 122, p. 871-882.

Dade, L.A., Gao, F.Q., Kovacevic, N., Roy, P., Rockel, C., O'Toole, C.M., Lobaugh, N.J., Feinstein, A., Levine, B. & Black, S.E., 2004. Semiautomatic brain region extraction: a method of parcellating brain regions from structural magnetic resonance images. *NeuroImage*, 22, p. 1492-1502.

Dale, A.M., Fischl, B. & Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage*, 9, p. 179-194.

Dalton, C.M., Brex, P.A., Jenkins, R., Fox, N.C., Miszkiel, K.A., Crum, W.R., O'Riordan, J.I., Plant, G.T., Thompson, A.J. & Miller, D.H., 2002a. Progressive ventricular enlargement in patients with clinically isolated syndromes is associated with the early development of multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 73, p. 141-147.

Dalton, C.M., Brex, P.A., Miszkiel, K.A., Hickman, S.J., MacManus, D.G., Plant, G.T., Thompson, A.J. & Miller, D.H., 2002b. Application of the new McDonald criteria to patients with clinically isolated syndromes suggestive of multiple sclerosis. *Annals of Neurology*, 52, p. 47-53.

Dalton, C.M., Chard, D.T., Davies, G.R., Miszkiel, K.A., Altmann, D.R., Fernando, K., Plant, G.T., Thompson, A.J. & Miller, D.H., 2004. Early development of multiple sclerosis is associated with progressive grey matter atrophy in patients presenting with clinically isolated syndromes. *Brain*, 127, p. 1101-1107.

Dalton, C.M., Miszkiel, K.A., O'Connor, P.W., Plant, G.T., Rice, G.P. & Miller, D.H., 2006. Ventricular enlargement in MS: one-year change at various stages of disease. *Neurology*, 66, p. 693-698.

Dastidar, P., Heinonen, T., Lehtimäki, T., Ukkonen, M., Peltola, J., Erilä, T., Laasonen, E. & Elovaara, I., 1999. Volumes of brain atrophy and plaques correlated with neurological disability in secondary progressive multiple sclerosis. *Journal of the Neurological Sciences*, 165, p. 36-42.

Datta, S., Sajja, B.R., He, R., Wolinsky, J.S., Gupta, R.K. & Narayana, P.A., 2006. Segmentation and quantification of black holes in multiple sclerosis. *NeuroImage*, 29, p. 467-474.

Davies, G.R., Ramió-Torrentà, L., Hadjiprocopis, A., Chard, D.T., Griffin, C.M., Rashid, W., Barker, G.J., Kapoor, R., Thompson, A.J. & Miller, D.H., 2004. Evidence for grey matter MTR abnormality in minimally disabled patients with early relapsing-remitting multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 75, p. 998-1002.

De Keyser, J., Zeinstra, E., Mostert, J. & Wilczak, N., 2004. β_2-adrenoceptor involvement in inflammatory demyelination and axonal degeneration in multiple sclerosis. *Trends in Pharmacological Sciences*, 25, p. 67-71.

De Stefano, N., Bartolozzi, M.L., Nacmias, B., Zipoli, V., Mortilla, M., Guidi, L., Siracusa, G., Sorbi, S., Federico, A. & Amato, M.P., 2004. Influence of apolipoprotein E ε4 genotype on brain tissue integrity in relapsing-remitting multiple sclerosis. *Archives of Neurology*, 61, p. 536-540.

De Stefano, N., Matthews, P.M., Filippi, M., Agosta, F., De Luca, M., Bartolozzi, M.L., Guidi, L., Ghezzi, A., Montanari, E., Cifelli, A., Federico, A. & Smith, S.M., 2003. Evidence of early cortical atrophy in MS: relevance to white matter changes and disability. *Neurology*, 60, p. 1157-1162.

De Stefano, N., Narayanan, S., Francis, G.S., Arnaoutelis, R., Tartaglia, M.C., Antel, J.P., Matthews, P.M. & Arnold, D.L., 2001. Evidence of axonal damage in the early stages of multiple sclerosis and its relevance to disability. *Archives of Neurology*, 58, p. 65-70.

Deloire, M.S., Salort, E., Bonnet, M., Arimone, Y., Boudineau, M., Amieva, H., Barroso, B., Ouallet, J.C., Pachai, C., Galliaud, E., Petry, K.G., Dousset, V., Fabrigoule, C. & Brochet, B., 2005. Cognitive impairment as marker of diffuse brain abnormalities in early relapsing remitting multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 76, p. 519-526.

Diestel, A., Aktas, O., Hackel, D., Häke, I., Meier, S., Raine, C.S., Nitsch, R., Zipp, F. & Ullrich, O., 2003. Activation of microglial poly(ADP-ribose)-polymerase-1 by cholesterol breakdown products during neuroinflammation: a link between demyelination and neuronal damage. *Journal of Experimental Medicine*, 198, p. 1729-1740.

Droogan, A.G., Clark, C.A., Werring, D.J., Barker, G.J., McDonald, W.I. & Miller, D.H., 1999. Comparison of multiple sclerosis clinical subgroups using navigated spin echo diffusion-weighted imaging. *Magnetic Resonance Imaging*, 17, p. 653-661.

Dutta, R., McDonough, J., Yin, X., Peterson, J., Chang, A., Torres, T., Gudz, T., Macklin, W.B., Lewis, D.A., Fox, R.J., Rudick, R., Mirnics, K. & Trapp, B.D., 2006. Mitochondrial dysfunction as a cause of axonal degeneration in multiple sclerosis patients. *Annals of Neurology*, 59, p. 478-489.

Enzinger, C., Ropele, S., Smith, S., Strasser-Fuchs, S., Poltrum, B., Schmidt, H., Matthews, P.M. & Fazekas, F., 2004. Accelerated evolution of brain atrophy and "black holes" in MS patients with APOE-ε4. *Annals of Neurology*, 55, p. 563-569.

Evangelou, N., Esiri, M.M., Smith, S., Palace, J. & Matthews, P.M., 2000a. Quantitative pathological evidence for axonal loss in normal appearing white matter in multiple sclerosis. *Annals of Neurology*, 47, p. 391-395.

Evangelou, N., Konz, D., Esiri, M.M., Smith, S., Palace, J. & Matthews, P.M., 2000b. Regional axonal loss in the corpus callosum correlates with cerebral white matter lesion volume and distribution in multiple sclerosis. *Brain*, 123, p. 1845-1849.

Fennema-Notestine, C., Ozyurt, I.B., Clark, C.P., Morris, S., Bischoff-Grethe, A., Bondi, M.W., Jernigan, T.L., Fischl, B., Segonne, F., Shattuck, D.W., Leahy, R.M., Rex, D.E., Toga, A.W., Zou, K.H. & Brown, G.G., 2006. Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: effects of diagnosis, bias correction, and slice location. *Human Brain Mapping*, 27, p. 99-113.

Fernando, K.T., McLean, M.A., Chard, D.T., MacManus, D.G., Dalton, C.M., Miszkiel, K.A., Gordon, R.M., Plant, G.T., Thompson, A.J. & Miller, D.H., 2004. Elevated white matter myo-inositol in clinically isolated syndromes suggestive of multiple sclerosis. *Brain*, 127, p. 1361-1369.

Fernando, K.T., Tozer, D.J., Miszkiel, K.A., Gordon, R.M., Swanton, J.K., Dalton, C.M., Barker, G.J., Plant, G.T., Thompson, A.J. & Miller, D.H., 2005. Magnetization transfer histograms in clinically isolated syndromes suggestive of multiple sclerosis. *Brain*, 128, p. 2911-2925.

Feuillet, L., Reuter, F., Audoin, B., Malikova, I., Barrau, K., Ali-Chèrif, A. & Pelletier, J., 2007. Early cognitive impairment in patients with clinically isolated syndrome suggestive of multiple sclerosis. *Multiple Sclerosis*, 13, p. 124-127.

Filippi, M., Bozzali, M., Rovaris, M., Gonen, O., Kesavadas, C., Ghezzi, A., Martinelli, V., Grossman, R.I., Scotti, G., Comi, G. & Falini, A., 2003. Evidence for widespread axonal damage at the earliest clinical stage of multiple sclerosis. *Brain*, 126, p. 433-437.

Filippi, M., Horsfield, M.A., Bressi, S., Martinelli, V., Baratti, C., Reganati, P., Campi, A., Miller, D.H. & Comi, G., 1995. Intra- and inter-observer agreement of brain MRI lesion volume measurements in multiple sclerosis. A comparison of techniques. *Brain*, 118, p. 1593-1600.

Filippi, M., Mastronardo, G., Rocca, M.A., Pereira, C. & Comi, G., 1998. Quantitative volumetric analysis of brain magnetic resonance imaging from patients with multiple sclerosis. *Journal of the Neurological Sciences*, 158, p. 148-153.

Filippi, M., Rovaris, M., Inglese, M., Barkhof, F., De Stefano, N., Smith, S. & Comi, G., 2004. Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet*, 364, p. 1489-1496.

Fischer, J.S., Rudick, R.A., Cutter, G.R. & Reingold, S.C., 1999. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. *Multiple Sclerosis*, 5, p. 244-250.

Fischl, B. & Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences U S A*, 97, p. 11050-11055.

Fisher, E., Cothren, R.M., Tkach, J.A., Masaryk, T.J. & Cornhill J.F., 1997. Knowledge-based 3D segmentation of MR images for quantitative MS lesion tracking. *SPIE Medical Imaging*, 3034, p. 599-610.

Fisher, E., Rudick, R.A., Cutter, G., Baier, M., Miller, D., Weinstock-Guttman, B., Mass, M.K., Dougherty, D.S. & Simonian, N.A., 2000. Relationship between brain atrophy and

disability: an 8-year follow-up study of multiple sclerosis patients. *Multiple Sclerosis*, 6, p. 373-377.

Fisher, E., Rudick, R.A., Simon, J.H., Cutter, G., Baier, M., Lee, J.C., Miller, D., Weinstock-Guttman, B., Mass, M.K., Dougherty, D.S. & Simonian, N.A., 2002. Eight-year follow-up study of brain atrophy in patients with MS. *Neurology*, 59, p. 1412-1420.

Forbes, K.P., Pipe, J.G., Bird, C.R. & Heiserman, J.E., 2001. PROPELLER MRI: clinical testing of a novel technique for quantification and compensation of head motion. *Journal of Magnetic Resonance Imaging*, 14, p. 215-222.

Forbes, R.B. & Swingler, R.J., 1999. Estimating the prevalence of multiple sclerosis in the United Kingdom by using capture-recapture methodology. *American Journal of Epidemiology*, 149, p. 1016-1024.

Fox, N.C., Black, R.S., Gilman, S., Rossor, M.N., Griffith, S.G., Jenkins, L. & Koller, M., 2005. Effects of Aβ immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. *Neurology*, 64, p. 1563-1572.

Fox, N.C., Cousens, S., Scahill, R., Harvey, R.J. & Rossor, M.N., 2000a. Using serial registered brain magnetic resonance imaging to measure disease progression in Alzheimer disease: power calculations and estimates of sample size to detect treatment effects. *Archives of Neurology*, 57, p. 339-344.

Fox, N.C. & Freeborough, P.A., 1997. Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease. *Journal of Magnetic Resonance Imaging*, 7, p. 1069-1075.

Fox, N.C., Jenkins, R., Leary, S.M., Stevenson, V.L., Losseff, N.A., Crum, W.R., Harvey, R.J., Rossor, M.N., Miller, D.H. & Thompson, A.J., 2000b. Progressive cerebral atrophy in MS: a serial study using registered, volumetric MRI. *Neurology*, 54, p. 807-812.

Frackowiak, R.S.J., Friston, K., Frith, C.D., Dolan, R.J. & Mazziotta, J.C., 1997. Human Brain Function. *Academic Press, USA.*

Frank, J.A., Richert, N., Bash, C., Stone, L., Calabresi, P.A., Lewis, B., Stone, R., Howard, T. & McFarland, H.F., 2004. Interferon-β-1b slows progression of atrophy in RRMS: Three-year follow-up in NAb- and NAb+ patients. *Neurology*, 62, p. 719-725.

Freeborough, P.A. & Fox, N.C., 1997. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Transactions on Medical Imaging*, 16, p. 623-629.

Freeborough, P.A. & Fox, N.C., 1998. Modeling brain deformations in Alzheimer disease by fluid registration of serial 3D MR images. *Journal of Computer Assisted Tomography*, 22, p. 838-843.

Freeborough, P.A., Fox, N.C. & Kitney, R.I., 1997. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine*, 53, p. 15-25.

Freeborough, P.A., Woods, R.P. & Fox, N.C., 1996. Accurate registration of serial 3D MR brain images and its application to visualizing change in neurodegenerative disorders. *Journal of Computer Assisted Tomography*, 20, p. 1012-1022.

Fritz, D.A., Dwyer, M.G., Bagnato, F., Watts, K.L., Bratina, A., Zorzon, M., Durastanti, V., Locatelli, L., Millefiorini, E. & Zivadinov, R., 2006. Effect of MRI coregistration on serial short-term brain volume changes in multiple sclerosis. *Neurological Research*, 28, p. 275-279.

Frost, C., Kenward, M.G. & Fox, N.C., 2004. The analysis of repeated 'direct' measures of change illustrated with an application in longitudinal imaging. *Statistics in Medicine*, 23, p. 3275-3286.

Gallo, A., Rovaris, M., Riva, R., Ghezzi, A., Benedetti, B., Martinelli, V., Falini, A., Comi, G. & Filippi, M., 2005. Diffusion-tensor magnetic resonance imaging detects normal-appearing white matter damage unrelated to short-term disease activity in patients at the earliest clinical stage of multiple sclerosis. *Archives of Neurology*, 62, p. 803-808.

Garaci, F.G., Colangelo, V., Ludovici, A., Gaudiello, F., Marziali, S., Centonze, D., Boffa, L., Simonetti, G. & Floris, R., 2007. A diffusion longitudinal MR imaging study in

normal-appearing white matter in untreated relapsing-remitting multiple sclerosis. *AJNR American Journal of Neuroradiology*, 28, p. 475-478.

Gasperini, C., Paolillo, A., Giugni, E., Galgani, S., Bagnato, F., Mainero, C., Onesti, E., Bastianello, S. & Pozzilli, C., 2002. MRI brain volume changes in relapsing-remitting multiple sclerosis patients treated with interferon beta-1a. *Multiple Sclerosis*, 8, p. 119-123.

Gasperini, C., Rovaris, M., Sormani, M.P., Bastianello, S., Pozzilli, C., Comi, G. & Filippi, M., 2001. Intra-observer, inter-observer and inter-scanner variations in brain MRI volume measurements in multiple sclerosis. *Multiple Sclerosis*, 7, p. 27-31.

Gawne-Cain, M.L., Webb, S., Tofts, P. & Miller, D.H., 1996. Lesion volume measurement in multiple sclerosis: how important is accurate repositioning? *Journal of Magnetic Resonance Imaging*, 6, p. 705-713.

Ge, Y., Grossman, R.I., Udupa, J.K., Babb, J.S., Nyúl, L.G. & Kolson, D.L., 2001. Brain atrophy in relapsing-remitting multiple sclerosis: fractional volumetric analysis of gray matter and white matter. *Radiology*, 220, p. 606-610.

Ge, Y., Grossman, R.I., Udupa, J.K., Fulton, J., Constantinescu, C.S., Gonzales-Scarano, F., Babb, J.S., Mannon, L.J., Kolson, D.L. & Cohen, J.A., 2000a. Glatiramer acetate (Copaxone) treatment in relapsing-remitting MS: quantitative MR assessment. *Neurology*, 54, p. 813-817.

Ge, Y., Grossman, R.I., Udupa, J.K., Wei, L., Mannon, L.J., Polansky, M. & Kolson, D.L., 2000b. Brain atrophy in relapsing-remitting multiple sclerosis and secondary progressive multiple sclerosis: longitudinal quantitative analysis. *Radiology*, 214, p. 665-670.

Geurts, J.J., Bø, L., Pouwels, P.J., Castelijns, J.A., Polman, C.H. & Barkhof, F., 2005. Cortical lesions in multiple sclerosis: combined postmortem MR imaging and histopathology. *AJNR American Journal of Neuroradiology*, 26, p. 572-577.

Geurts, J.J., Bø, L., Roosendaal, S.D., Hazes, T., Daniëls, R., Barkhof, F., Witter, M.P., Huitinga, I., van der Valk, P., 2007. Extensive hippocampal demyelination in multiple sclerosis. *Journal of Neuropathology and Experimental Neurology*, 66, p. 819-827.

Geurts, J.J., Reuling, I.E., Vrenken, H., Uitdehaag, B.M., Polman, C.H., Castelijns, J.A., Barkhof, F. & Pouwels, P.J., 2006. MR spectroscopic evidence for thalamic and hippocampal, but not cortical, damage in multiple sclerosis. *Magnetic Resonance in Medicine*, 55, p. 478-483.

Ghezzi, A., Martinelli, V., Torri, V., Zaffaroni, M., Rodegher, M., Comi, G., Zibetti, A. & Canal, N., 1999. Long-term follow-up of isolated optic neuritis: the risk of developing multiple sclerosis, its outcome, and the prognostic role of paraclinical tests. *Journal of Neurology*, 246, p. 770-775.

Gispert, J.D., Reig, S., Pascau, J., Vaquero, J.J., García-Barreno, P. & Desco, M., 2004. Method for bias field correction of brain T1-weighted magnetic resonance images minimizing segmentation error. *Human Brain Mapping*, 22, p. 133-144.

Goldszal, A.F., Davatzikos, C., Pham, D.L., Yan, M.X., Bryan, R.N. & Resnick, S.M., 1998. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *Journal of Computer Assisted Tomography*, 22, p. 827-837.

Gonen, O., Catalaa, I., Babb, J.S., Ge, Y., Mannon, L.J., Kolson, D.L. & Grossman, R.I., 2000. Total brain N-acetylaspartate: a new measure of disease load in MS. *Neurology*, 54, p. 15-19.

Grimaud, J., Lai, M., Thorpe, J., Adeleine, P., Wang, L., Barker, G.J., Plummer, D.L., Tofts, P.S., McDonald, W.I. & Miller, D.H., 1996. Quantification of MRI lesion load in multiple sclerosis: a comparison of three computer-assisted techniques. *Magnetic Resonance Imaging*, 14, p. 495-505.

Gunter, J.L., Shiung, M.M., Manduca, A. & Jack, C.R., Jr., 2003. Methodological considerations for measuring rates of brain atrophy. *Journal of Magnetic Resonance Imaging*, 18, p. 16-24.

Hahn, H.K., Jolly, B., Lee, M., Krastel, D., Rexilius, J., Drexl, J., Schlüter, M., Terwey, B. & Peitgen, H.O., 2004. How accurate is brain volumetry? A methodological evaluation. *Lecture Notes In Computer Science*, 3216, p. 335-342.

Han, C., 1969. Testing the homogeneity of variances in a two-way classification. *Biometrics*, 25, p. 153-158.

Han, X., Jovicich, J., Salat, D., van der Kouwe, A., Quinn, B., Czanner, S., Busa, E., Pacheco, J., Albert, M., Killiany, R., Maguire, P., Rosas, D., Makris, N., Dale, A., Dickerson, B. & Fischl, B., 2006. Reliability of MRI-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer. *NeuroImage*, 32, p. 180-194.

Hardmeier, M., Wagenpfeil, S., Freitag, P., Fisher, E., Rudick, R.A., Kooijmans, M., Clanet, M., Radue, E.W. & Kappos, L., 2005. Rate of brain atrophy in relapsing MS decreases during treatment with IFNbeta-1a. *Neurology*, 64, p. 236-240.

Hardmeier, M., Wagenpfeil, S., Freitag, P., Fisher, E., Rudick, R.A., Kooijmans-Coutinho, M., Clanet, M., Radue, E.W. & Kappos, L., 2003. Atrophy is detectable within a 3-month period in untreated patients with active relapsing remitting multiple sclerosis. *Archives of Neurology*, 60, p. 1736-1739.

Heinonen, T., Dastidar, P., Kauppinen, P., Malmivuo, J. & Eskola, H., 1998. Semi-automatic tool for segmentation and volumetric analysis of medical images. *Medical and Biological Engineering and Computing*, 36, p. 291-296.

Heinonen, T., Eskola, H., Dastidar, P., Laarne, P. & Malmivuo, J., 1997. Segmentation of T1 MR scans for reconstruction of resistive head models. *Computer Methods and Programs in Biomedicine*, 54, p. 173-181.

Held, U., Heigenhauser, L., Shang, C., Kappos, L. & Polman, C., 2005. Predictors of relapse rate in MS clinical trials. *Neurology*, 65, p. 1769-1773.

Henley, S.M., Frost, C., MacManus, D.G., Warner, T.T., Fox, N.C. & Tabrizi, S.J., 2006. Increased rate of whole-brain atrophy over 6 months in early Huntington disease. *Neurology*, 67, p. 694-696.

Hobart, J., Kalkers, N., Barkhof, F., Uitdehaag, B., Polman, C. & Thompson, A., 2004. Outcome measures for multiple sclerosis clinical trials: relative measurement precision of the Expanded Disability Status Scale and Multiple Sclerosis Functional Composite. *Multiple Sclerosis*, 10, p. 41-46.

Holmes, C.J., Hoge, R., Collins, L., Woods, R., Toga, A.W. & Evans, A.C., 1998. Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*, 22, p. 324-333.

Hoogervorst, E.L., Polman, C.H. & Barkhof, F., 2002. Cerebral volume changes in multiple sclerosis patients treated with high-dose intravenous methylprednisolone. *Multiple Sclerosis*, 8, p. 415-419.

Horakova, D., Cox, J.L., Havrdova, E., Hussein, S., Dolezal, O., Cookfair, D., Dwyer, M.G., Seidl, Z., Bergsland, N., Vaneckova, M. & Zivadinov, R., 2007. Evolution of different MRI measures in patients with active relapsing-remitting multiple sclerosis over 2 and 5 years. A case control study. *Journal of Neurology Neurosurgery and Psychiatry*, published online 5 Jun 2007, doi:10.1136/jnnp.2007.120378.

Horsfield, M.A., Rovaris, M., Rocca, M.A., Rossi, P., Benedict, R.H., Filippi, M. & Bakshi, R., 2003. Whole-brain atrophy in multiple sclerosis measured by two segmentation processes from various MRI sequences. *Journal of the Neurological Sciences*, 216, p. 169-177.

Ingle, G.T., Stevenson, V.L., Miller, D.H., Leary, S.M., Rovaris, M., Barkhof, F., Brochet, B., Dousset, V., Filippi, M., Montalban, X., Kalkers, N.F., Polman, C.H., Rovira, A. & Thompson, A.J., 2002. Two-year follow-up study of primary and transitional progressive multiple sclerosis. *Multiple Sclerosis*, 8, p. 108-114.

Ingle, G.T., Stevenson, V.L., Miller, D.H. & Thompson, A.J., 2003. Primary progressive multiple sclerosis: a 5-year clinical and MR study. *Brain*, 126, p. 2528-2536.

Inglese, M., Mancardi, G.L., Pagani, E., Rocca, M.A., Murialdo, A., Saccardi, R., Comi, G. & Filippi, M., 2004. Brain tissue loss occurs after suppression of enhancement in patients with multiple sclerosis treated with autologous haematopoietic stem cell transplantation. *Journal of Neurology Neurosurgery and Psychiatry*, 75, p. 643-644.

Jacobs, L.D., Beck, R.W., Simon, J.H., Kinkel, R.P., Brownscheidle, C.M., Murray, T.J., Simonian, N.A., Slasor, P.J. & Sandrock, A.W., 2000. Intramuscular interferon beta-1a therapy initiated during a first demyelinating event in multiple sclerosis. *New England Journal of Medicine*, 343, p. 898-904.

Jacobs, L.D., Cookfair, D.L., Rudick, R.A., Herndon, R.M., Richert, J.R., Salazar, A.M., Fischer, J.S., Goodkin, D.E., Granger, C.V., Simon, J.H., Alam, J.J., Bartoszak, D.M., Bourdette, D.N., Braiman, J., Brownscheidle, C.M., Coats, M.E., Cohan, S.L., Dougherty, D.S., Kinkel, R.P., Mass, M.K., Munschauer, F.E., Priore, R.L., Pullicino, P.M., Scherokman, B.J., Weinstock-Guttman, B., Whitman, R.H. & The Multiple Sclerosis Collaborative Research Group, 1996. Intramuscular interferon beta-1 alpha for disease progression in relapsing multiple sclerosis. *Annals of Neurology*, 39, p. 285-294.

Jasperse, B., Minneboo, A., de Groot, V., Kalkers, N.F., van Helden, P.E., Uitdehaag, B.M., Barkhof, F. & Polman, C.H., 2007a. Determinants of cerebral atrophy rate at the time of diagnosis of multiple sclerosis. *Archives of Neurology*, 64, p. 190-194.

Jasperse, B., Valsasina, P., Neacsu, V., Knol, D.L., De Stefano, N., Enzinger, C., Smith, S.M., Ropele, S., Korteweg, T., Giorgio, A., Anderson, V., Polman, C.H., Filippi, M., Miller, D.H., Rovaris, M., Barkhof, F. & Vrenken, H., 2007b. Intercenter agreement of brain atrophy measurement in multiple sclerosis patients using manually-edited SIENA and SIENAX. *Journal of Magnetic Resonance Imaging*, 26, p. 881-885.

Jasperse, B., Vrenken, H., Sanz-Arigita, E., de Groot, V., Smith, S.M., Polman, C.H. & Barkhof, F., 2007c. Regional brain atrophy development is related to specific aspects of clinical dysfunction in multiple sclerosis. *NeuroImage*, 38, p. 529-537.

Jenkinson, M. & Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5, p. 143-156.

Johnson, K.P., Brooks, B.R., Cohen, J.A., Ford, C.C., Goldstein, J., Lisak, R.P., Myers, L.W., Panitch, H.S., Rose, J.W. & Schiffer, R.B., Vollmer, T., Weiner, L.P., Wolinsky, J.S. & The Copolymer 1 Multiple Sclerosis Study Group, 1995. Copolymer 1 reduces relapse rate and improves disability in relapsing-remitting multiple sclerosis: results of a phase III multicenter, double-blind placebo-controlled trial. *Neurology*, 45, p. 1268-1276.

Johnson, K.P., Brooks, B.R., Cohen, J.A., Ford, C.C., Goldstein, J., Lisak, R.P., Myers, L.W., Panitch, H.S., Rose, J.W., Schiffer, R.B., Vollmer, T., Weiner, L.P. & Wolinsky, J.S., 1998. Extended use of glatiramer acetate (Copaxone) is well tolerated and maintains its clinical effect on multiple sclerosis relapse rate and degree of disability. Copolymer 1 Multiple Sclerosis Study Group. *Neurology*, 50, p. 701-708.

Jovicich, J., Czanner, S., Greve, D., Haley, E., van der Kouwe, A., Gollub, R., Kennedy, D., Schmitt, F., Brown, G., Macfall, J., Fischl, B. & Dale, A., 2006. Reliability in multi-site structural MRI studies: effects of gradient non-linearity correction on phantom and human data. *NeuroImage*, 30, p. 436-443.

Kalkers, N.F., Ameziane, N., Bot, J.C., Minneboo, A., Polman, C.H. & Barkhof, F., 2002. Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. *Archives of Neurology*, 59, p. 1572-1576.

Kalkers, N.F., Bergers, E., Castelijns, J.A., van Walderveen, M.A., Bot, J.C., Adèr, H.J., Polman, C.H. & Barkhof, F., 2001a. Optimizing the association between disability and biological markers in MS. *Neurology*, 57, p. 1253-1258.

Kalkers, N.F., Bergers, L., de Groot, V., Lazeron, R.H., van Walderveen, M.A., Uitdehaag, B.M., Polman, C.H. & Barkhof, F., 2001b. Concurrent validity of the MS Functional Composite using MRI as a biological disease marker. *Neurology*, 56, p. 215-219.

Kangarlu, A., Bourekas, E.C., Ray-Chaudhury, A. & Rammohan, K.W., 2007. Cerebral cortical lesions in multiple sclerosis detected by MR imaging at 8 Tesla. *AJNR American Journal of Neuroradiology*, 28, p. 262-266.

Kappos, L., Freedman, M.S., Polman, C.H., Edan, G., Hartung, H.P., Miller, D.H., Montalban, X., Barkhof, F., Radü, E.W., Bauer, L., Dahms, S., Lanius, V., Pohl, C. & Sandbrink, R., 2007. Effect of early versus delayed interferon beta-1b treatment on disability after a first clinical event suggestive of multiple sclerosis: a 3-year follow-up analysis of the BENEFIT study. *Lancet*, 370, p. 389-397.

Kappos, L., Moeri, D., Radue, E.W., Schoetzau, A., Schweikert, K., Barkhof, F., Miller, D., Guttmann, C.R., Weiner, H.L., Gasperini, C. & Filippi, M., 1999. Predictive value of

gadolinium-enhanced magnetic resonance imaging for relapse rate and changes in disability or impairment in multiple sclerosis: a meta-analysis. *Lancet*, 353, p. 964-969.

Kassubek, J., Tumani, H., Ecker, D., Kurt, A., Ludolph, A.C. & Juengling, F.D., 2003. Age-related brain parenchymal fraction is significantly decreased in young multiple sclerosis patients: a quantitative MRI study. *NeuroReport*, 14, p. 427-430.

Kidd, D., Barkhof, F., McConnell, R., Algra, P.R., Allen, I.V. & Revesz, T., 1999. Cortical lesions in multiple sclerosis. *Brain*, 122, p. 17-26.

Kim, J.S., Singh, V., Lee, J.K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., Lee, J.M., Kim, S.I. & Evans, A.C., 2005. Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *NeuroImage*, 27, p. 210-221.

Kornek, B., Storch, M.K., Bauer, J., Djamshidian, A., Weissert, R., Wallstroem, E., Stefferl, A., Zimprich, F., Olsson, T., Linington, C., Schmidbauer, M. & Lassmann, H., 2001. Distribution of a calcium channel subunit in dystrophic axons in multiple sclerosis and experimental autoimmune encephalomyelitis. *Brain*, 124, p. 1114-1124.

Kovacevic, N., Lobaugh, N.J., Bronskill, M.J., Levine, B., Feinstein, A. & Black, S.E., 2002. A robust method for extraction and automatic segmentation of brain images. *NeuroImage*, 17, p. 1087-1100.

Kuhlmann, T., Lingfeld, G., Bitsch, A., Schuchardt, J. & Brück, W., 2002. Acute axonal damage in multiple sclerosis is most extensive in early disease stages and decreases over time. *Brain*, 125, p. 2202-2212.

Kurtzke, J.F., 1983. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, 33, p. 1444-1452.

Kutzelnigg, A., Lucchinetti, C.F., Stadelmann, C., Brück, W., Rauschka, H., Bergmann, M., Schmidbauer, M., Parisi, J.E. & Lassmann, H., 2005. Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain*, 128, p. 2705-2712.

Lachin, J.M., 2004. The role of measurement reliability in clinical trials. *Clinical Trials*, 1, p. 553-566.

Langer-Gould, A., Popat, R.A., Huang, S.M., Cobb, K., Fontoura, P., Gould, M.K. & Nelson, L.M., 2006. Clinical and demographic predictors of long-term disability in patients with relapsing-remitting multiple sclerosis: a systematic review. *Archives of Neurology*, 63, p. 1686-1691.

Lappe-Siefke, C., Goebbels, S., Gravel, M., Nicksch, E., Lee, J., Braun, P.E., Griffiths, I.R. & Nave, K.A., 2003. Disruption of Cnp1 uncouples oligodendroglial functions in axonal support and myelination. *Nature Genetics*, 33, p. 366-374.

Lazeron, R.H., Boringa, J.B., Schouten, M., Uitdehaag, B.M., Bergers, E., Lindeboom, J., Eikelenboom, M.I., Scheltens, P.H., Barkhof, F. & Polman, C.H., 2005. Brain atrophy and lesion load as explaining parameters for cognitive impairment in multiple sclerosis. *Multiple Sclerosis*, 11, p. 524-531.

Lazeron, R.H., de Sonneville, L.M., Scheltens, P., Polman, C.H. & Barkhof, F., 2006. Cognitive slowing in multiple sclerosis is strongly associated with brain volume reduction. *Multiple Sclerosis*, 12, p. 760-768.

Leech, S., Kirk, J., Plumb, J. & McQuaid, S., 2007. Persistent endothelial abnormalities and blood-brain barrier leak in primary and secondary progressive multiple sclerosis. *Neuropathology and Applied Neurobiology*, 33, p. 86-98.

Leigh, R., Ostuni, J., Pham, D., Goldszal, A., Lewis, B.K., Howard, T., Richert, N., McFarland, H. & Frank, J.A., 2002. Estimating cerebral atrophy in multiple sclerosis patients from various MR pulse sequences. *Multiple Sclerosis*, 8, p. 420-429.

Levin, L.I., Munger, K.L., Rubertone, M.V., Peck, C.A., Lennette, E.T., Spiegelman, D. & Ascherio, A., 2005. Temporal relationship between elevation of Epstein-Barr virus antibody titers and initial onset of neurological symptoms in multiple sclerosis. *Journal of the American Medical Association*, 293, p. 2496-2500.

Lewis, E.B. & Fox, N.C., 2004. Correction of differential intensity inhomogeneity in longitudinal MR images. *NeuroImage*, 23, p. 75-83.

Lin, X. & Blumhardt, L.D., 2001. Inflammation and atrophy in multiple sclerosis: MRI associations with disease course. *Journal of the Neurological Sciences*, 189, p. 99-104.

Lin, X., Blumhardt, L.D., & Constantinescu, C.S., 2003. The relationship of brain and cervical cord volume to disability in clinical subtypes of multiple sclerosis: a three-dimensional MRI study. *Acta Neurologica Scandinavica*, 108, p. 401-406.

Lin, X., Tench, C.R., Morgan, P.S. & Constantinescu, C.S., 2007. Use of combined conventional and quantitative MRI to quantify pathology related to cognitive impairment in multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, published online 7 Aug 2007, doi:10.1136/jnnp.2006.112177.

Lincoln, M.R., Montpetit, A., Cader, M.Z., Saarela, J., Dyment, D.A., Tiislar, M., Ferretti, V., Tienari, P.J., Sadovnick, A.D., Peltonen, L., Ebers, G.C. & Hudson, T.J., 2005. A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nature Genetics*, 37, p. 1108-1112.

Liu, C., Edwards, S., Gong, Q., Roberts, N. & Blumhardt, L.D., 1999. Three dimensional MRI estimates of brain and spinal cord atrophy in multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 66, p. 323-330.

Locatelli, L., Zivadinov, R., Grop, A. & Zorzon, M., 2004. Frontal parenchymal atrophy measures in multiple sclerosis. *Multiple Sclerosis*, 10, p. 562-568.

Losseff, N.A., Wang, L., Lai, H.M., Yoo, D.S., Gawne-Cain, M.L., McDonald, W.I., Miller, D.H. & Thompson, A.J., 1996. Progressive cerebral atrophy in multiple sclerosis. A serial MRI study. *Brain*, 119, p. 2009-2019.

Lu, F., Selak, M., O'Connor, J., Croul, S., Lorenzana, C., Butunoi, C. & Kalman, B., 2000. Oxidative damage to mitochondrial DNA and activity of mitochondrial enzymes in chronic active lesions of multiple sclerosis. *Journal of the Neurological Sciences*, 177, p. 95-103.

Lublin, F.D. & Reingold, S.C., 1996. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory

Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology*, 46, p. 907-911.

Lucchinetti, C., Brück, W., Parisi, J., Scheithauer, B., Rodriguez, M. & Lassmann, H., 2000. Heterogeneity of multiple sclerosis lesions: implications for the pathogenesis of demyelination. *Annals of Neurology*, 47, p. 707-717.

Lukas, C., Hahn, H.K., Bellenberg, B., Rexilius, J., Schmid, G., Schimrigk, S.K., Przuntek, H., Köster, O. & Peitgen, H.O., 2004. Sensitivity and reproducibility of a new fast 3D segmentation technique for clinical MR-based brain volumetry in multiple sclerosis. *Neuroradiology*, 46, p. 906-915.

Luks, T.L., Goodkin, D.E., Nelson, S.J., Majumdar, S., Bacchetti, P., Portnoy, D. & Sloan, R., 2000. A longitudinal study of ventricular volume in early relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, 6, p. 332-337.

Luo, J., Zhu, Y., Clarysse, P. & Magnin, I., 2005. Correction of bias field in MR images using singularity function analysis. *IEEE Transactions on Medical Imaging*, 24, p. 1067-1085.

Lycklama á Nijeholt, G.J., van Walderveen, M.A., Castelijns, J.A., van Waesberghe, J.H., Polman, C., Scheltens, P., Rosier, P.F., Jongen, P.J. & Barkhof, F., 1998. Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms. *Brain*, 121, p. 687-697.

Madabhushi, A. & Udupa, J.K., 2005. Interplay between intensity standardization and inhomogeneity correction in MR image processing. *IEEE Transactions on Medical Imaging*, 24, p. 561-576.

Maes, F., Collignon, A., Vandermeulen, D., Marchal, G. & Suetens, P., 1997. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16, p. 187-198.

Martola, J., Stawiarz, L., Fredrikson, S., Hillert, J., Bergström, J., Flodmark, O. & Kristoffersen, W.M., 2007. Progression of non-age-related callosal brain atrophy in

multiple sclerosis: a 9-year longitudinal MRI study representing four decades of disease development. *Journal of Neurology Neurosurgery and Psychiatry*, 78, p. 375-380.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R. & Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 356, p. 1293-1322.

McDonald, W.I., Compston, A., Edan, G., Goodkin, D., Hartung, H.P., Lublin, F.D., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., van den Noort, S., Weinshenker, B.Y. & Wolinsky, J.S., 2001. Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, 50, p. 121-127.

Miller, D.H., Soon, D., Fernando, K.T., MacManus, D.G., Barker, G.J., Yousry, T.A., Fisher, E., O'Connor, P.W., Phillips, J.T., Polman, C.H., Kappos, L., Hutchinson, M., Havrdova, E., Lublin, F.D., Giovannoni, G., Wajgt, A., Rudick, R., Lynn, F., Panzara, M.A. & Sandrock, A.W., 2007. MRI outcomes in a placebo-controlled trial of natalizumab in relapsing MS. *Neurology*, 68, p. 1390-1401.

Mitchell, J.R., Karlik, S.J., Lee, D.H., Eliasziw, M., Rice, G.P. & Fenster, A., 1996. The variability of manual and computer assisted quantification of multiple sclerosis lesion volumes. *Medical Physics*, 23, p. 85-97.

Möller, A., Wiedemann, G., Rohde, U., Backmund, H. & Sonntag, A., 1994. Correlates of cognitive impairment and depressive mood disorder in multiple sclerosis. *Acta Psychiatrica Scandinavica*, 89, p. 117-121.

Molyneux, P.D., Kappos, L., Polman, C., Pozzilli, C., Barkhof, F., Filippi, M., Yousry, T., Hahn, D., Wagner, K., Ghazi, M., Beckmann, K., Dahlke, F., Losseff, N., Barker, G.J., Thompson, A.J. & Miller, D.H., 2000a. The effect of interferon beta-1b treatment on MRI measures of cerebral atrophy in secondary progressive multiple sclerosis. *Brain*, 123, p. 2256-2263.

Molyneux, P.D., Miller, D.H., Filippi, M., Yousry, T., Kappos, L., Gasperini, C., Adèr, H.J. & Barkhof, F., 2000b. The use of magnetic resonance imaging in multiple sclerosis treatment trials: power calculations for annual lesion load measurement. *Journal of Neurology*, 247, p. 34-40.

Molyneux, P.D., Miller, D.H., Filippi, M., Yousry, T.A., Radu, E.W., Adèr, H.J. & Barkhof, F., 1999. Visual analysis of serial T2-weighted MRI in multiple sclerosis: intra- and interobserver reproducibility. *Neuroradiology*, 41, p. 882-888.

Molyneux, P.D., Tofts, P.S., Fletcher, A., Gunn, B., Robinson, P., Gallagher, H., Moseley, I.F., Barker, G.J. & Miller, D.H., 1998a. Precision and reliability for measurement of change in MRI lesion volume in multiple sclerosis: a comparison of two computer assisted techniques. *Journal of Neurology Neurosurgery and Psychiatry*, 65, p. 42-47.

Molyneux, P.D., Tubridy, N., Parker, G.J., Barker, G.J., MacManus, D.G., Tofts, P.S., Moseley, I.F. & Miller, D.H., 1998b. The effect of section thickness on MR lesion detection and quantification in multiple sclerosis. *AJNR American Journal of Neuroradiology*, 19, p. 1715-1720.

Molyneux, P.D., Wang, L., Lai, M., Barker, G.J., Tofts, P.S., Moseley, I.F. & Miller, D.H., 1998c. Quantitative techniques for lesion load measurement in multiple sclerosis: an assessment of the global threshold technique after non uniformity and histogram matching corrections. *European Journal of Neurology*, 5, p. 55-60.

Morgen, K., Sammer, G., Courtney, S.M., Wolters, T., Melchior, H., Blecker, C.R., Oschmann, P., Kaps, M. & Vaitl, D., 2006. Evidence for a direct association between cortical atrophy and cognitive impairment in relapsing-remitting MS. *NeuroImage*, 30, p. 891-898.

Morrissey, S.P., Miller, D.H., Kendall, B.E., Kingsley, D.P., Kelly, M.A., Francis, D.A., MacManus, D.G. & McDonald, W.I., 1993. The significance of brain magnetic resonance imaging abnormalities at presentation with clinically isolated syndromes suggestive of multiple sclerosis. A 5-year follow-up study. *Brain*, 116, p. 135-146.

Mumford, C.J., Fraser, M.B., Wood, N.W. & Compston, D.A., 1992. Multiple sclerosis in the Cambridge health district of East Anglia. *Journal of Neurology Neurosurgery and Psychiatry*, 55, p. 877-882.

Nelson, H.E., 1982. The National Adult Reading Test. 2nd Edition. NFER-Nelson, Windsor, UK.

Noseworthy, J.H., 1994. Clinical scoring methods for multiple sclerosis. *Annals of Neurology*, 36, S80-S85.

Oh, J., Henry, R.G., Genain, C., Nelson, S.J. & Pelletier, D., 2004a. Mechanisms of normal appearing corpus callosum injury related to pericallosal T1 lesions in multiple sclerosis using directional diffusion tensor and 1H MRS imaging. *Journal of Neurology Neurosurgery and Psychiatry*, 75, p. 1281-1286.

Oh, J., Pelletier, D. & Nelson, S.J., 2004b. Corpus callosum axonal injury in multiple sclerosis measured by proton magnetic resonance spectroscopic imaging. *Archives of Neurology*, 61, p. 1081-1086.

Olivares, T., Nieto, A., Sánchez, M.P., Wollmann, T., Hernández, M.A. & Barroso, J., 2005. Pattern of neuropsychological impairment in the early phase of relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, 11, p. 191-197.

Optic Neuritis Study Group, 1997. The 5-year risk of MS after optic neuritis. Experience of the optic neuritis treatment trial. *Neurology*, 49, p. 1404-1413.

Oreja-Guevara, C., Rovaris, M., Iannucci, G., Valsasina, P., Caputo, D., Cavarretta, R., Sormani, M.P., Ferrante, P., Comi, G. & Filippi, M., 2005. Progressive gray matter damage in patients with relapsing-remitting multiple sclerosis: a longitudinal diffusion tensor magnetic resonance imaging study. *Archives of Neurology*, 62, p. 578-584.

Pagani, E., Filippi, M., Rocca, M.A. & Horsfield, M.A., 2005a. A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis. *NeuroImage*, 26, p. 258-265.

Pagani, E., Rocca, M.A., Gallo, A., Rovaris, M., Martinelli, V., Comi, G. & Filippi, M., 2005b. Regional brain atrophy evolves differently in patients with multiple sclerosis according to clinical phenotype. *AJNR American Journal of Neuroradiology*, 26, p. 341-346.

Paolillo, A., Piattella, M.C., Pantano, P., Di Legge, S., Caramia, F., Russo, P., Lenzi, G.L. & Pozzilli, C., 2004. The relationship between inflammation and atrophy in clinically isolated syndromes suggestive of multiple sclerosis. A monthly MRI study after triple-dose gadolinium-DTPA. *Journal of Neurology*, 251, p. 432-439.

Paolillo, A., Pozzilli, C., Gasperini, C., Giugni, E., Mainero, C., Giuliani, S., Tomassini, V., Millefiorini, E. & Bastianello, S., 2000. Brain atrophy in relapsing-remitting multiple sclerosis: relationship with 'black holes', disease duration and clinical disability. *Journal of the Neurological Sciences*, 174, p. 85-91.

Paolillo, A., Pozzilli, C., Giugni, E., Tomassini, V., Gasperini, C., Fiorelli, M., Mainero, C., Horsfield, M., Galgani, S., Bastianello, S. & Buttinelli, C., 2002. A 6-year clinical and MRI follow-up study of patients with relapsing-remitting multiple sclerosis treated with Interferon-beta. *European Journal of Neurology*, 9, p. 645-655.

Pekmezovic, T., Drulovic, J., Milenkovic, M., Jarebinski, M., Stojsavljevic, N., Mesaros, S., Kisic, D. & Kostic, J., 2006. Lifestyle factors and multiple sclerosis: A case-control study in Belgrade. *Neuroepidemiology*, 27, p. 212-216.

Pelletier, J., Suchet, L., Witjas, T., Habib, M., Guttmann, C.R., Salamon, G., Lyon-Caen, O. & Ali-Chérif, A., 2001. A longitudinal study of callosal atrophy and interhemispheric dysfunction in relapsing-remitting multiple sclerosis. *Archives of Neurology*, 58, p. 105-111.

Pestalozza, I.F., Pozzilli, C., Di Legge, S., Piattella, M.C., Pantano, P., Caramia, F., Pasqualetti, P. & Lenzi, G.L., 2005. Monthly brain magnetic resonance imaging scans in patients with clinically isolated syndrome. *Multiple Sclerosis*, 11, p. 390-394.

Peterson, J.W., Bø, L., Mørk, S., Chang, A. & Trapp, B.D., 2001. Transected neurites, apoptotic neurons, and reduced inflammation in cortical multiple sclerosis lesions. *Annals of Neurology*, 50, p. 389-400.

Pham, D.L. & Prince, J.L., 1999. Adaptive fuzzy segmentation of magnetic resonance images. *IEEE Transactions on Medical Imaging*, 18, p. 737-752.

Pipe, J.G., 1999. Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging. *Magnetic Resonance in Medicine*, 42, p. 963-969.

Piras, M.R., Magnano, I., Canu, E.D., Paulus, K.S., Satta, W.M., Soddu, A., Conti, M., Achene, A., Solinas, G. & Aiello, I., 2003. Longitudinal study of cognitive dysfunction in multiple sclerosis: neuropsychological, neuroradiological, and neurophysiological findings. *Journal of Neurology Neurosurgery and Psychiatry*, 74, p. 878-885.

Pittock, S.J., McClelland, R.L., Mayr, W.T., Jorgensen, N.W., Weinshenker, B.G., Noseworthy, J. & Rodriguez, M., 2004. Clinical implications of benign multiple sclerosis: a 20-year population-based follow-up study. *Annals of Neurology*, 56, p. 303-306.

Plummer, D.L., 1992. DispImage: a display and analysis tool for medical images. *Revista di Neuroradiologia*, 5, p. 489-495.

Polman, C.H., O'Connor, P.W., Havrdova, E., Hutchinson, M., Kappos, L., Miller, D.H., Phillips, J.T., Lublin, F.D., Giovannoni, G., Wajgt, A., Toal, M., Lynn, F., Panzara, M.A. & Sandrock, A.W., 2006. A randomized, placebo-controlled trial of natalizumab for relapsing multiple sclerosis. *New England Journal of Medicine*, 354, p. 899-910.

Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., Lublin, F.D., Metz, L.M., McFarland, H.F., O'Connor, P.W., Sandberg-Wollheim, M., Thompson, A.J., Weinshenker, B.G. & Wolinsky, J.S., 2005. Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals of Neurology*, 58, p. 840-846.

Portaccio, E., Amato, M.P., Bartolozzi, M.L., Zipoli, V., Mortilla, M., Guidi, L., Siracusa, G., Sorbi, S., Federico, A. & De Stefano, N., 2006. Neocortical volume decrease in relapsing-remitting multiple sclerosis with mild cognitive impairment. *Journal of the Neurological Sciences*, 245, p. 195-199.

Poser, C.M., Paty, D.W., Scheinberg, L., McDonald, W.I., Davis, F.A., Ebers, G.C., Johnson, K.P., Sibley, W.A., Silberberg, D.H. & Tourtellotte, W.W., 1983. New

diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Annals of Neurology*, 13, p. 227-231.

Preboske, G.M., Gunter, J.L., Ward, C.P. & Jack, C.R. Jr., 2006. Common MRI acquisition non-idealities significantly impact the output of the boundary shift integral method of measuring brain atrophy on serial MRI. *NeuroImage*, 30, p. 1196-1202.

Price, R.R., Axel, L., Morgan, T., Newman, R., Perman, W., Schneiders, N., Selikson, M., Wood, M.L. & Thomas, S.R., 1990. Quality assurance methods and phantoms for magnetic resonance imaging. *Medical Physics*, 17, p. 287-295.

Prinster, A., Quarantelli, M., Orefice, G., Lanzillo, R., Brunetti, A., Mollica, C., Salvatore, E., Morra, V.B., Coppola, G., Vacca, G., Alfano, B. & Salvatore, M., 2006. Grey matter loss in relapsing-remitting multiple sclerosis: a voxel-based morphometry study. *NeuroImage*, 29, p. 859-867.

Quarantelli, M., Ciarmiello, A., Morra, V.B., Orefice, G., Larobina, M., Lanzillo, R., Schiavone, V., Salvatore, E., Alfano, B. & Brunetti, A., 2003. Brain tissue volume changes in relapsing-remitting multiple sclerosis: correlation with lesion load. *NeuroImage*, 18, p. 360-366.

Ranjeva, J.P., Pelletier, J., Confort-Gouny, S., Ibarrola, D., Audoin, B., Le Fur, Y., Viout, P., Ali Chérif, A. & Cozzone, P.J., 2003. MRI/MRS of corpus callosum in patients with clinically isolated syndrome suggestive of multiple sclerosis. *Multiple Sclerosis*, 9, p. 554-565.

Rao, A.B., Richert, N., Howard, T., Lewis, B.K., Bash, C.N., McFarland, H.F. & Frank, J.A., 2002. Methylprednisolone effect on brain volume and enhancing lesions in MS before and during IFNβ-1b. *Neurology*, 59, p. 688-694.

Rao, S.M., Glatt, S., Hammeke, T.A., McQuillen, M.P., Khatri, B.O., Rhodes, A.M. & Pollard, S., 1985. Chronic progressive multiple sclerosis. Relationship between cerebral ventricular size and neuropsychological impairment. *Archives of Neurology*, 42, p. 678-682.

Rao, S.M., Leo, G.J., Bernardin, L. & Unverzagt, F., 1991. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*, 41, p. 685-691.

Rao, S.M., Leo, G.J., Haughton, V.M., Aubin-Faubert, P. & Bernardin, L., 1989. Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. *Neurology*, 39, p. 161-166.

Rashid, W., Davies, G.R., Chard, D.T., Griffin, C.M., Altmann, D.R., Thompson, A.J. & Miller, D.H., 2007. Relationship of triple dose contrast enhanced lesions with clinical measures and brain atrophy in early relapsing-remitting multiple sclerosis: a two-year longitudinal study. *Multiple Sclerosis*, 13, p. 178-185.

Rashid, W., Hadjiprocopis, A., Griffin, C.M., Chard, D.T., Davies, G.R., Barker, G.J., Tofts, P.S., Thompson, A.J. & Miller, D.H., 2004. Diffusion tensor imaging of early relapsing-remitting multiple sclerosis with histogram analysis using automated segmentation and brain volume correction. *Multiple Sclerosis*, 10, p. 9-15.

Redmond, I.T., Barbosa, S., Blumhardt, L.D. & Roberts, N., 2000. Short-term ventricular volume changes on serial MRI in multiple sclerosis. *Acta Neurologica Scandinavica*, 102, p. 99-105.

Revesz, T., Kidd, D., Thompson, A.J., Barnard, R.O. & McDonald, W.I., 1994. A comparison of the pathology of primary and secondary progressive multiple sclerosis. *Brain*, 117, p. 759-765.

Rey, D., Subsol, G., Delingette, H. & Ayache, N., 2002. Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. *Medical Image Analysis*, 6, p. 163-179.

Richert, N.D., Howard, T., Frank, J.A., Stone, R., Ostuni, J., Ohayon, J., Bash, C. & McFarland, H.F., 2006. Relationship between inflammatory lesions and cerebral atrophy in multiple sclerosis. *Neurology*, 66, p. 551-556.

Rocca, M.A., Pagani, E., Ghezzi, A., Falini, A., Zaffaroni, M., Colombo, B., Scotti, G., Comi, G. & Filippi, M., 2003. Functional cortical changes in patients with multiple

sclerosis and nonspecific findings on conventional magnetic resonance imaging scans of the brain. *NeuroImage*, 19, p. 826-836.

Roosendaal, S.D., Moraal, B., Vrenken, H., Castelijns, J.A., Pouwels, P.J., Barkhof, F., Geurts, J.J., 2008. In vivo MR imaging of hippocampal lesions in multiple sclerosis. *Journal of Magnetic Resonance Imaging*, published online 26 Feb 2008, doi 10.1002/jmri.21294

Rovaris, M., Agosta, F., Sormani, M.P., Inglese, M., Martinelli, V., Comi, G. & Filippi, M., 2003. Conventional and magnetization transfer MRI predictors of clinical multiple sclerosis evolution: a medium-term follow-up study. *Brain*, 126, p. 2323-2332.

Rovaris, M., Comi, G., Rocca, M.A., Valsasina, P., Ladkani, D., Pieri, E., Weiss, S., Shifroni, G., Wolinsky, J.S. & Filippi, M., 2007. Long-term follow-up of patients treated with glatiramer acetate: a multicentre, multinational extension of the European/Canadian double-blind, placebo-controlled, MRI-monitored trial. *Multiple Sclerosis*, 13, p. 502-508.

Rovaris, M., Comi, G., Rocca, M.A., Wolinsky, J.S. & Filippi, M., 2001. Short-term brain volume change in relapsing-remitting multiple sclerosis: effect of glatiramer acetate and implications. *Brain*, 124, p. 1803-1812.

Rovaris, M., Filippi, M., Falautano, M., Minicucci, L., Rocca, M.A., Martinelli, V. & Comi, G., 1998. Relation between MR abnormalities and patterns of cognitive impairment in multiple sclerosis. *Neurology*, 50, p. 1601-1608.

Rovaris, M., Gallo, A., Valsasina, P., Benedetti, B., Caputo, D., Ghezzi, A., Montanari, E., Sormani, M.P., Bertolotto, A., Mancardi, G., Bergamaschi, R., Martinelli, V., Comi, G. & Filippi, M., 2005a. Short-term accrual of gray matter pathology in patients with progressive multiple sclerosis: an in vivo study using diffusion tensor MRI. *NeuroImage*, 24, p. 1139-1146.

Rovaris, M., Gambini, A., Gallo, A., Falini, A., Ghezzi, A., Benedetti, B., Sormani, M.P., Martinelli, V., Comi, G. & Filippi, M., 2005b. Axonal injury in early multiple sclerosis is irreversible and independent of the short-term disease evolution. *Neurology*, 65, p. 1626-1630.

Rovaris, M., Inglese, M., van Schijndel, R.A., Sormani, M.P., Rodegher, M., Comi, G. & Filippi, M., 2000. Sensitivity and reproducibility of volume change measurements of different brain portions on magnetic resonance imaging in patients with multiple sclerosis. *Journal of Neurology*, 247, p. 960-965.

Rudick, R.A., Cutter, G., Baier, M., Fisher, E., Dougherty, D., Weinstock-Guttman, B., Mass, M.K., Miller, D. & Simonian, N.A., 2001. Use of the Multiple Sclerosis Functional Composite to predict disability in relapsing MS. *Neurology*, 56, p. 1324-1330.

Rudick, R.A., Fisher, E., Lee, J.C., Duda, J.T. & Simon, J., 2000. Brain atrophy in relapsing multiple sclerosis: relationship to relapses, EDSS, and treatment with interferon beta-1a. *Multiple Sclerosis*, 6, p. 365-372.

Rudick, R.A., Fisher, E., Lee, J.C., Simon, J. & Jacobs, L., 1999. Use of the brain parenchymal fraction to measure whole brain atrophy in relapsing-remitting MS. *Neurology*, 53, p. 1698-1704.

Rudick, R.A., Lee, J.C., Simon, J. & Fisher, E., 2006. Significance of T2 lesions in multiple sclerosis: A 13-year longitudinal study. *Annals of Neurology*, 60, p. 236-242.

Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O. & Hawkes, D.J., 1999. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18, p. 712-721.

Sahakian, B.J. & Owen, A.M., 1992. Computerized assessment in neuropsychiatry using CANTAB: discussion paper. *Journal of the Royal Society of Medicine*, 85, p. 399-402.

Sailer, M., Fischl, B., Salat, D., Tempelmann, C., Schönfeld, M.A., Busa, E., Bodammer, N., Heinze, H.J. & Dale, A., 2003. Focal thinning of the cerebral cortex in multiple sclerosis. *Brain*, 126, p. 1734-1744.

Sailer, M., Losseff, N.A., Wang, L., Gawne-Cain, M.L., Thompson, A.J. & Miller, D.H., 2001. T1 lesion load and cerebral atrophy as a marker for clinical progression in patients with multiple sclerosis. A prospective 18 months follow-up study. *European Journal of Neurology*, 8, p. 37-42.

Saindane, A.M., Ge, Y., Udupa, J.K., Babb, J.S., Mannon, L.J. & Grossman, R.I., 2000. The effect of gadolinium-enhancing lesions on whole brain atrophy in relapsing-remitting MS. *Neurology*, 55, p. 61-65.

Salat, D.H., Buckner, R.L., Snyder, A.Z., Greve, D.N., Desikan, R.S., Busa, E., Morris, J.C., Dale, A.M. & Fischl, B., 2004. Thinning of the cerebral cortex in aging. *Cerebral Cortex*, 14, p. 721-730.

Sampson, H., 1956. Pacing and performance on a serial addition task. *Canadian Journal of Psychology*, 10, p. 1219-1225.

Sánchez, I., Hassinger, L., Paskevich, P.A., Shine, H.D. & Nixon, R.A., 1996. Oligodendroglia regulate the regional expansion of axon caliber and local accumulation of neurofilaments during development independently of myelin formation. *Journal of Neuroscience*, 16, p. 5095-5105.

Sanfilipo, M.P., Benedict, R.H., Sharma, J., Weinstock-Guttman, B. & Bakshi, R., 2005. The relationship between whole brain volume and disability in multiple sclerosis: a comparison of normalized gray vs. white matter with misclassification correction. *NeuroImage*, 26, p. 1068-1077.

Sanfilipo, M.P., Benedict, R.H., Weinstock-Guttman, B. & Bakshi, R., 2006. Gray and white matter brain atrophy and neuropsychological impairment in multiple sclerosis. *Neurology*, 66, p. 685-692.

Sastre-Garriga, J., Ingle, G.T., Chard, D.T., Cercignani, M., Ramió-Torrentà, L., Miller, D.H. & Thompson, A.J., 2005a. Grey and white matter volume changes in early primary progressive multiple sclerosis: a longitudinal study. *Brain*, 128, p. 1454-1460.

Sastre-Garriga, J., Ingle, G.T., Chard, D.T., Ramió-Torrentà, L., McLean, M.A., Miller, D.H. & Thompson, A.J., 2005b. Metabolite changes in normal-appearing gray and white matter are linked with disability in early primary progressive multiple sclerosis. *Archives of Neurology*, 62, p. 569-573.

Sastre-Garriga, J., Ingle, G.T., Chard, D.T., Ramió-Torrentà, L., Miller, D.H. & Thompson, A.J., 2004. Grey and white matter atrophy in early clinical stages of primary progressive multiple sclerosis. *NeuroImage*, 22, p. 353-359.

Sayao, A.L., Devonshire, V. & Tremlett, H., 2007. Longitudinal follow-up of "benign" multiple sclerosis at 20 years. *Neurology*, 68, p. 496-500.

Scahill, R.I., Frost, C., Jenkins, R., Whitwell, J.L., Rossor, M.N. & Fox, N.C., 2003. A longitudinal study of brain volume changes in normal aging using serial registered magnetic resonance imaging. *Archives of Neurology*, 60, p. 989-994.

Scahill, R.I., Schott, J.M., Stevens, J.M., Rossor, M.N. & Fox, N.C., 2002. Mapping the evolution of regional atrophy in Alzheimer's disease: unbiased analysis of fluid-registered serial MRI. *Proceedings of the National Academy of Sciences U S A*, 99, p. 4703-4707.

Schmierer, K., Scaravilli, F., Altmann, D.R., Barker, G.J. & Miller, D.H., 2004. Magnetization transfer ratio and myelin in postmortem multiple sclerosis brain. *Annals of Neurology*, 56, p. 407-415.

Schott, J.M., Price, S.L., Frost, C., Whitwell, J.L., Rossor, M.N. & Fox, N.C., 2005. Measuring atrophy in Alzheimer disease: a serial MRI study over 6 and 12 months. *Neurology*, 65, p. 119-124.

Schwid, S.R., Goodman, A.D., Weinstein, A., McDermott, M.P. & Johnson, K.P., 2007. Cognitive function in relapsing multiple sclerosis: minimal changes in a 10-year clinical trial. *Journal of the Neurological Sciences*, 255, p. 57-63.

Sepulcre, J., Sastre-Garriga, J., Cercignani, M., Ingle, G.T., Miller, D.H. & Thompson, A.J., 2006. Regional gray matter atrophy in early primary progressive multiple sclerosis: a voxel-based morphometry study. *Archives of Neurology*, 63, p. 1175-1180.

Sharma, J., Sanfilipo, M.P., Benedict, R.H., Weinstock-Guttman, B., Munschauer, F.E. III & Bakshi, R., 2004. Whole-brain atrophy in multiple sclerosis measured by automated versus semiautomated MR imaging segmentation. *AJNR American Journal of Neuroradiology*, 25, p. 985-996.

Shattuck, D.W., Sandor-Leahy, S.R., Schaper, K.A., Rottenberg, D.A. & Leahy, R.M., 2001. Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13, p. 856-876.

Shen, D. & Davatzikos, C., 2003. Very high-resolution morphometry using mass-preserving deformations and HAMMER elastic registration. *NeuroImage*, 18, p. 28-41.

Simon, J.H., Jacobs, L.D., Campion, M.K., Rudick, R.A., Cookfair, D.L., Herndon, R.M., Richert, J.R., Salazar, A.M., Fischer, J.S., Goodkin, D.E., Simonian, N., Lajaunie, M., Miller, D.E., Wende, K., Martens-Davidson, A., Kinkel, R.P., Munschauer, F.E. III & Brownscheidle, C.M., 1999. A longitudinal study of brain atrophy in relapsing multiple sclerosis. *Neurology*, 53, p. 139-148.

Sled, J.G., Zijdenbos, A.P. & Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging*, 17, p. 87-97.

Smith, A., 1982. Symbol Digit Modalities Test. Los Angeles: Western Psychological Services.

Smith, K.J., Kapoor, R., Hall, S.M. & Davies, M., 2001a. Electrically active axons degenerate when exposed to nitric oxide. *Annals of Neurology*, 49, p. 470-476.

Smith, S.M., 2002. Fast robust automated brain extraction. *Human Brain Mapping*, 17, p. 143-155.

Smith, S.M., De Stefano, N., Jenkinson, M. & Matthews, P.M., 2001b. Normalized accurate measurement of longitudinal brain change. *Journal of Computer Assisted Tomography*, 25, p. 466-475.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M. & Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, S208-S219.

Smith, S.M., Rao, A., De Stefano, N., Jenkinson, M., Schott, J.M., Matthews, P.M. & Fox, N.C., 2007. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: Cross-validation of BSI, SIENA and SIENAX. *NeuroImage*, 36, p. 1200-1206.

Smith, S.M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P.M., Federico, A. & De Stefano, N., 2002. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage*, 17, p. 479-489.

Snaith, R.P. & Zigmond, A.S., 1986. The hospital anxiety and depression scale. *British Medical Journal*, 292, p. 344-344.

Soon, D., Tozer, D.J., Altmann, D.R., Tofts, P.S. & Miller, D.H., 2007. Quantification of subtle blood-brain barrier disruption in non-enhancing lesions in multiple sclerosis: a study of disease and lesion subtypes. *Multiple Sclerosis*, 13, p. 884-894.

Sormani, M.P., Miller, D.H., Comi, G., Barkhof, F., Rovaris, M., Bruzzi, P. & Filippi, M., 2001. Clinical trials of multiple sclerosis monitored with enhanced MRI: new sample size calculations based on large data sets. *Journal of Neurology Neurosurgery and Psychiatry*, 70, p. 494-499.

Sormani, M.P., Rovaris, M., Valsasina, P., Wolinsky, J.S., Comi, G. & Filippi, M., 2004. Measurement error of two different techniques for brain atrophy assessment in multiple sclerosis. *Neurology*, 62, p. 1432-1434.

Stevenson, V.L., Ingle, G.T., Miller, D.H. & Thompson, A.J., 2004. Magnetic resonance imaging predictors of disability in primary progressive multiple sclerosis: a 5-year study. *Multiple Sclerosis*, 10, p. 398-401.

Stevenson, V.L., Miller, D.H., Leary, S.M., Rovaris, M., Barkhof, F., Brochet, B., Dousset, V., Filippi, M., Hintzen, R., Montalban, X., Polman, C.H., Rovira, A., de Sa, J. & Thompson, A.J., 2000. One year follow up study of primary and transitional progressive multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 68, p. 713-718.

Stevenson, V.L., Smith, S.M., Matthews, P.M., Miller, D.H. & Thompson, A.J., 2002. Monitoring disease activity and progression in primary progressive multiple sclerosis

using MRI: sub-voxel registration to identify lesion changes and to detect cerebral atrophy. *Journal of Neurology*, 249, p. 171-177.

Studholme, C., Hill, D.L. & Hawkes, D.J., 1999. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32, p. 71-86.

Tan, I.L., van Schijndel, R.A., Fazekas, F., Filippi, M., Freitag, P., Miller, D.H., Yousry, T.A., Pouwels, P.J., Adàr, H.J. & Barkhof, F., 2002a. Image registration and subtraction to detect active T2 lesions in MS: an interobserver study. *Journal of Neurology*, 249, p. 767-773.

Tan, I.L., van Schijndel, R.A., Pouwels, P.J., Adàr, H.J. & Barkhof, F., 2002b. Serial isotropic three-dimensional fast FLAIR imaging: using image registration and subtraction to reveal active multiple sclerosis lesions. *AJR American Journal of Roentgenology*, 179, p. 777-782.

Taylor, I., Butzkueven, H., Litewka, L., MacGregor, L.R., Szoeke, C., Cook, M., Mitchell, P., Kilpatrick, T.J. & Tubridy, N., 2004. Serial MRI in multiple sclerosis: a prospective pilot study of lesion load, whole brain volume and thalamic atrophy. *Journal of Clinical Neuroscience*, 11, p. 153-158.

Tekok-Kilic, A., Benedict, R.H., Weinstock-Guttman, B., Dwyer, M.G., Carone, D., Srinivasaraghavan, B., Yella, V., Abdelrahman, N., Munschauer, F., Bakshi, R. & Zivadinov, R., 2007. Independent contributions of cortical gray matter atrophy and ventricle enlargement for predicting neuropsychological impairment in multiple sclerosis. *NeuroImage*, 36, p. 1294-1300.

Thacker, N.A., Jackson, A., Moriarty, D. & Vokurka, E., 1999. Improved quality of re-sliced MR images using re-normalized sinc interpolation. *Journal of Magnetic Resonance Imaging*, 10, p. 582-588.

The IFNB Multiple Sclerosis Study Group, 1995. Interferon beta-1b in the treatment of multiple sclerosis - final outcome of the randomized controlled trial. *Neurology*, 45, p. 1277-1285.

Thompson, A.J., Kermode, A.G., Wicks, D., MacManus, D.G., Kendall, B.E., Kingsley, D.P. & McDonald, W.I., 1991. Major differences in the dynamics of primary and secondary progressive multiple sclerosis. *Annals of Neurology*, 29, p. 53-62.

Tiberio, M., Chard, D.T., Altmann, D.R., Davies, G., Griffin, C.M., McLean, M.A., Rashid, W., Sastre-Garriga, J., Thompson, A.J. & Miller, D.H., 2006. Metabolite changes in early relapsing-remitting multiple sclerosis. A two year follow-up study. *Journal of Neurology*, 253, p. 224-230.

Tiberio, M., Chard, D.T., Altmann, D.R., Davies, G., Griffin, C.M., Rashid, W., Sastre-Garriga, J., Thompson, A.J. & Miller, D.H., 2005. Gray and white matter volume changes in early RRMS: a 2-year longitudinal study. *Neurology*, 64, p. 1001-1007.

Tintoré, M., Rovira, A., Rio, J., Nos, C., Grivé, E., Téllez, N., Pelayo, R., Comabella, M. & Montalban, X., 2005. Is optic neuritis more benign than other first attacks in multiple sclerosis? *Annals of Neurology*, 57, p. 210-215.

Tintoré, M., Rovira, A., Río, J., Nos, C., Grivé, E., Téllez, N., Pelayo, R., Comabella, M., Sastre-Garriga, J. & Montalban, X., 2006. Baseline MRI predicts future attacks and disability in clinically isolated syndromes. *Neurology*, 67, p. 968-972.

Traboulsee, A., Dehmeshki, J., Peters, K.R., Griffin, C.M., Brex, P.A., Silver, N., Ciccarrelli, O., Chard, D.T., Barker, G.J., Thompson, A.J. & Miller, D.H., 2003. Disability in multiple sclerosis is related to normal appearing brain tissue MTR histogram abnormalities. *Multiple Sclerosis*, 9, p. 566-573.

Trapp, B.D., Peterson, J., Ransohoff, R.M., Rudick, R., Mørk, S. & Bø, L., 1998. Axonal transection in the lesions of multiple sclerosis. *New England Journal of Medicine*, 338, p. 278-285.

Tubridy, N., Ader, H.J., Barkhof, F., Thompson, A.J. & Miller, D.H., 1998. Exploratory treatment trials in multiple sclerosis using MRI: sample size calculations for relapsing-remitting and secondary progressive subgroups using placebo controlled parallel groups. *Journal of Neurology Neurosurgery and Psychiatry*, 64, p. 50-55.

Turner, B., Lin, X., Calmon, G., Roberts, N. & Blumhardt, L.D., 2003. Cerebral atrophy and disability in relapsing-remitting and secondary progressive multiple sclerosis over four years. *Multiple Sclerosis*, 9, p. 21-27.

Turner, B., Ramli, N., Blumhardt, L.D. & Jaspan, T., 2001. Ventricular enlargement in multiple sclerosis: a comparison of three-dimensional and linear MRI estimates. *Neuroradiology*, 43, p. 608-614.

Udupa, J.K. & Samarasekera, S., 1996. Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation. *Graphical Models and Image Processing*, 58, p. 246-261.

Udupa, J.K., Wei, L., Samarasekera, S., Miki, Y., van Buchem, M.A. & Grossman, R.I., 1997. Multiple sclerosis lesion quantification using fuzzy-connectedness principles. *IEEE Transactions on Medical Imaging*, 16, p. 598-609.

Ukkonen, M., Dastidar, P., Heinonen, T., Laasonen, E. & Elovaara, I., 2003. Volumetric quantitation by MRI in primary progressive multiple sclerosis: volumes of plaques and atrophy correlated with neurological disability. *European Journal of Neurology*, 10, p. 663-669.

Valsasina, P., Benedetti, B., Rovaris, M., Sormani, M.P., Comi, G. & Filippi, M., 2005. Evidence for progressive gray matter loss in patients with relapsing-remitting MS. *Neurology*, 65, p. 1126-1128.

van der Mei, I.A., Ponsonby, A.L., Dwyer, T., Blizzard, L., Simmons, R., Taylor, B.V., Butzkueven, H. & Kilpatrick, T., 2003. Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case-control study. *British Medical Journal*, 327, p. 316-316.

Van Leemput, K., Maes, F., Vandermeulen, D. & Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18, p. 897-908.

van Waesberghe, J.H., Kamphorst, W., De Groot, C.J., van Walderveen, M.A., Castelijns, J.A., Ravid, R., Lycklama à Nijeholt, G.J., van der Valk, P., Polman, C.H., Thompson,

A.J. & Barkhof, F., 1999. Axonal loss in multiple sclerosis lesions: magnetic resonance imaging insights into substrates of disability. *Annals of Neurology*, 46, p. 747-754.

van Walderveen, M.A., Kamphorst, W., Scheltens, P., van Waesberghe, J.H., Ravid, R., Valk, J., Polman, C.H. & Barkhof, F., 1998. Histopathologic correlate of hypointense lesions on T1-weighted spin-echo MRI in multiple sclerosis. *Neurology*, 50, p. 1282-1288.

Vokurka, E.A., Thacker, N.A. & Jackson, A., 1999. A fast model independent method for automatic correction of intensity nonuniformity in MRI data. *Journal of Magnetic Resonance Imaging*, 10, p. 550-562.

Vovk, U., Pernuš, F. & Likar, B., 2004. MRI intensity inhomogeneity correction by combining intensity and spatial information. *Physics in Medicine and Biology*, 49, p. 4119-4133.

Vrenken, H., Barkhof, F., Uitdehaag, B.M., Castelijns, J.A., Polman, C.H. & Pouwels, P.J., 2005. MR spectroscopic evidence for glial increase but not for neuro-axonal damage in MS normal-appearing white matter. *Magnetic Resonance in Medicine*, 53, p. 256-266.

Vrenken, H., Geurts, J.J., Knol, D.L., Polman, C.H., Castelijns, J.A., Pouwels, P.J. & Barkhof, F., 2006a. Normal-appearing white matter changes vary with distance to lesions in multiple sclerosis. *AJNR American Journal of Neuroradiology*, 27, p. 2005-2011.

Vrenken, H., Pouwels, P.J., Geurts, J.J., Knol, D.L., Polman, C.H., Barkhof, F. & Castelijns, J.A., 2006b. Altered diffusion tensor in multiple sclerosis normal-appearing brain tissue: cortical diffusion changes seem related to clinical deterioration. *Journal of Magnetic Resonance Imaging*, 23, p. 628-636.

Vukusic, S. & Confavreux, C., 2003. Prognostic factors for progression of disability in the secondary progressive phase of multiple sclerosis. *Journal of the Neurological Sciences*, 206, p. 135-137.

Wechsler, D.A., 1981. Wechsler Adult Intelligence Test-Revised. New York: The Psychological Corporation, Harcourt Brace Jovanovich.

Whitwell, J.L., Crum, W.R., Watt, H.C. & Fox, N.C., 2001. Normalization of cerebral volumes by use of intracranial volume: implications for longitudinal quantitative MR imaging. *AJNR American Journal of Neuroradiology*, 22, p. 1483-1489.

Wicks, D.A., Tofts, P.S., Miller, D.H., du Boulay, G.H., Feinstein, A., Sacares, R.P., Harvey, I., Brenner, R. & McDonald, W.I., 1992. Volume measurement of multiple sclerosis lesions with magnetic resonance images. A preliminary study. *Neuroradiology*, 34, p. 475-479.

Wilkins, A., Majed, H., Layfield, R., Compston, A. & Chandran, S., 2003. Oligodendrocytes promote neuronal survival and axonal length by distinct intracellular mechanisms: a novel role for oligodendrocyte-derived glial cell line-derived neurotrophic factor. *Journal of Neuroscience*, 23, p. 4967-4974.

Williams, E.S. & McKeran, R.O., 1986. Prevalence of multiple sclerosis in a south London borough. *British Medical Journal*, 293, p. 237-239.

Wolinsky, J.S., Narayana, P.A., Johnson, K.P. & the Copolymer 1 Multiple Sclerosis Study Group, 2001. United States open-label glatiramer acetate extension trial for relapsing multiple sclerosis: MRI and clinical correlates. *Multiple Sclerosis*, 7, p. 33-41.

Woods, R.P., Grafton, S.T., Holmes, C.J., Cherry, S.R. & Mazziotta, J.C., 1998. Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22, p. 139-152.

Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C.S. & Aizenstein, H.J., 2007a. Optimum template selection for atlas-based segmentation. *NeuroImage*, 34, p. 1612-1618.

Wu, X., Kuusisto, H., Dastidar, P., Huhtala, H., Nikkari, S.T., Ukkonen, M., Höyhtyä, M. & Elovaara, I., 2007b. Once-weekly 22mug subcutaneous IFN-β-1a in secondary progressive MS: a 3-year follow-up study on brain MRI measurements and serum MMP-9 levels. *Acta Neurologica Scandinavica*, 116, p. 43-48.

Wu, Y., Warfield, S.K., Tan, I.L., Wells, W.M. III, Meier, D.S., van Schijndel, R.A., Barkhof, F. & Guttmann, C.R., 2006. Automated segmentation of multiple sclerosis lesion subtypes with multichannel MRI. *NeuroImage*, 32, p. 1205-1215.

Wylezinska, M., Cifelli, A., Jezzard, P., Palace, J., Alecci, M. & Matthews, P.M., 2003. Thalamic neurodegeneration in relapsing-remitting multiple sclerosis. *Neurology*, 60, p. 1949-1954.

Yao, S.Y., Stratton, C.W., Mitchell, W.M. & Sriram, S., 2001. CSF oligoclonal bands in MS include antibodies against Chlamydophila antigens. *Neurology*, 56, p. 1168-1176.

Zakrzewska-Pniewska, B., Styczynska, M., Podlecka, A., Samocka, R., Peplonska, B., Barcikowska, M. & Kwiecinski, H., 2004. Association of apolipoprotein E and myeloperoxidase genotypes to clinical course of familial and sporadic multiple sclerosis. *Multiple Sclerosis*, 10, p. 266-271.

Zhang, Y., Brady, M. & Smith, S., 2001a. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20, p. 45-57.

Zhang, Y., Goldszal, A., Butman, J. & Choyke, P., 2001b. Improving image contrast using principal component analysis for subsequent image segmentation. *Journal of Computer Assisted Tomography*, 25, p. 817-822.

Zivadinov, R., Bagnato, F., Nasuelli, D., Bastianello, S., Bratina, A., Locatelli, L., Watts, K., Finamore, L., Grop, A., Dwyer, M., Catalan, M., Clemenzi, A., Millefiorini, E., Bakshi, R. & Zorzon, M., 2004a. Short-term brain atrophy changes in relapsing-remitting multiple sclerosis. *Journal of the Neurological Sciences*, 223, p. 185-193.

Zivadinov, R., De Masi, R., Nasuelli, D., Bragadin, L.M., Ukmar, M., Pozzi-Mucelli, R.S., Grop, A., Cazzato, G. & Zorzon, M., 2001a. MRI techniques and cognitive impairment in the early phase of relapsing-remitting multiple sclerosis. *Neuroradiology*, 43, p. 272-278.

Zivadinov, R., Dwyer, M. & Watts, K., 2004b. Measurement of cerebral grey and white matter atrophy from various MRI pulse sequences using different segmentation algorithms. *Journal of Neurology*, 251, p. 89-89.

Zivadinov, R., Grop, A., Sharma, J., Bratina, A., Tjoa, C.W., Dwyer, M. & Zorzon, M., 2005. Reproducibility and accuracy of quantitative magnetic resonance imaging

techniques of whole-brain atrophy measurement in multiple sclerosis. *Journal of Neuroimaging*, 15, p. 27-36.

Zivadinov, R., Locatelli, L., Cookfair, D., Srinivasaraghavan, B., Bertolotto, A., Ukmar, M., Bratina, A., Maggiore, C., Bosco, A., Grop, A., Catalan, M. & Zorzon, M., 2007. Interferon beta-1a slows progression of brain atrophy in relapsing-remitting multiple sclerosis predominantly by reducing gray matter atrophy. *Multiple Sclerosis*, 13, p. 490-501.

Zivadinov, R., Locatelli, L., Stival, B., Bratina, A., Grop, A., Nasuelli, D., Brnabic-Razmilic, O. & Zorzon, M., 2003. Normalized regional brain atrophy measurements in multiple sclerosis. *Neuroradiology*, 45, p. 793-798.

Zivadinov, R., Sepcic, J., Nasuelli, D., De Masi, R., Bragadin, L.M., Tommasi, M.A., Zambito-Marsala, S., Moretti, R., Bratina, A., Ukmar, M., Pozzi-Mucelli, R.S., Grop, A., Cazzato, G. & Zorzon, M., 2001b. A longitudinal study of brain atrophy and cognitive disturbances in the early phase of relapsing-remitting multiple sclerosis. *Journal of Neurology Neurosurgery and Psychiatry*, 70, p. 773-780.

Zwemmer, J.N., van Veen, T., van Winsen, L., van Kamp, G.J., Barkhof, F., Polman, C.H. & Uitdehaag, B.M., 2004. No major association of ApoE genotype with disease characteristics and MRI findings in multiple sclerosis. *Multiple Sclerosis*, 10, p. 272-277.