Brain parenchymal fraction in healthy individuals and in clinical follow-up of multiple sclerosis

Mattias Vågberg
“Essentially, all models are wrong, but some are useful”

George E. P. Box

Till Maja och Edvin
Contents of Paper I and the PRISMA 2009 checklist are reproduced under a Creative Commons attribution license.
Contents of Paper II are bound by copyright to Elsevier and are reproduced with their permission.
# Table of Contents

Table of Contents i  
Abstract iv  
Original papers v  
Abbreviations vi  
Enkel sammanfattning på svenska viii  
Introduction 1  
  Multiple Sclerosis 1  
  Epidemiology and clinical course 1  
  The use of MRI in MS 3  
  Analysis of CSF in MS 3  
  The diagnosis of MS 4  
  Expanded Disability Status Scale (EDSS) 6  
  Clinical follow-up of MS 6  
Magnetic Resonance Imaging 7  
  The concept of MRI, explained briefly 7  
  The hydrogen atom and the MRI scanner 7  
  Precession 9  
  Recording a signal 9  
  Relaxation 10  
Limitations of conventional MRI 12  
Synthetic Tissue Mapping 12  
Brain atrophy 14  
Brain atrophy in MS 14  
The cause of MS-associated brain atrophy 15  
Quantification of brain atrophy 16  
Brain Parenchymal Fraction (BPF) 16  
Issues associated with the quantification of brain atrophy 18  
CSF analyses in MS 19  
Neurofilament light (NFL) 19  
Glial fibrillary acidic protein (GFAP) 19  
The relationships of NFL and GFAP to age 20  
The clinical care program for MS at Umeå University Hospital, Sweden. 20  
Rationale of this dissertation 20  
BPF and levels of NFL and GFAP in CSF in healthy individuals 20  
Comparing different methods for BPF determination 20  
Can the prognostic value of BPF be reproduced in clinical follow-up? 21  
Aims of this dissertation 21  
Materials and methods 22  
  Ethical statement 22  
  Study populations 23
Healthy volunteers (Papers I, II and IV) 23
Individuals with MS (Paper IV) 25
Definition of BPF and Percent brain volume change (PBVC) 28
MRI 28
Conventional $T_1$, $T_2$ weighted and FLAIR images (Paper II) 28
Synthetic Tissue Mapping (SyMap) (Paper I, II and IV) 28
Statistical Parametric Mapping (SPM) (Paper II) 29
Voxel-Based Morphometry (VBM) (Paper II) 29
3DSlicer (Paper II) 29
Lumbar Puncture and analysis of CSF 30
Lumbar puncture (Paper I) 30
Analysis of the levels of NFL in CSF samples (Paper I) 30
Analysis of level of GFAP in CSF (Paper I) 30
Systematic Review 30
Statistics 33

Results 35
Paper I 35
Levels of NFL in CSF 35
Levels of GFAP in CSF 35
BPF 35
Partial correlations 35
Paper II 36
Paper III 38
Results of the literature search 38
BPF in relation to age 39
BPF in relation to method 39
Paper IV 39
BPF in MS in relation to lesion count, MS type and disease duration as well as in relation to BPF in healthy individuals 39
 Associations with baseline EDSS and prediction of EDSS worsening 41

Discussion 44
Summary of the main results 44
The relationships between NFL and GFAP levels and BPF 44
BPF in relation to age 45
BPF in relation to post processing method 46
BPF in the MS population and its relationship with EDSS 47
Predicting EDSS worsening 47
Limitations of the studies 48
Paper I 48
Paper II 49
Paper III 50
Paper IV 51
How should the clinical follow-up of MS be designed? 53
Quality of Life (QoL) 53
Clinical disability assessment 55
Assessing inflammation and neuronal damage 55
  Relapses and MRI lesions 55
  Atrophy of the CNS 56
Measurement variation 57
Difficulties in interpreting the evidence from follow-up studies 57
Summarising the evidence for establishing a follow-up regime 58
The need for future research 60
Conclusions 61
Acknowledgements 62
References 64
Abstract

Background Multiple sclerosis (MS) is an autoimmune disease characterised by inflammatory damage to the central nervous system (CNS). Accumulated CNS injury can be quantified as brain atrophy, definable as a reduction in brain parenchymal fraction (BPF). BPF correlate with disability in MS and is used routinely as an endpoint in clinical trials. In 2009/2010, a new MS clinical care program, that includes follow-up of BPF, was introduced at Umeå University Hospital (NUS). Levels of neurofilament light polypeptide (NFL) and glial fibrillary acidic protein (GFAP) in cerebrospinal fluid (CSF) are markers of axonal and astrocytic injury, respectively, and also potential surrogate biomarkers for BPF decline. The goals of this thesis were to establish age-adjusted values of BPF in healthy individuals and to relate these to the BPF values from individuals with MS as well as to the levels of NFL and GFAP in CSF. Another goal was to investigate if expanded disability status scale (EDSS)-worsening could be predicted in a clinical MS cohort and if BPF measurements could contribute to such predictions. Methods A group of 111 healthy individuals volunteered to participate in the studies. A total of 106 of these underwent MRI with BPF measurements, 53 underwent lumbar puncture (LP) with measurement of NFL and GFAP and 48 underwent both MRI and LP. Three different automatic and one manual method were utilised to determine BPF. A literature search on BPF in healthy individuals was performed for the purpose of a systematic review. For studying disability progression in MS, all individuals with MS followed at NUS and included in the Swedish MS registry were included if they had matched data on BPF, EDSS and lesion load as part of clinical follow-up (n=278). Results BPF as well as NFL and GFAP levels in CSF were all associated with age. NFL was associated with BPF and GFAP, but only the association with GFAP was retained when adjusting for age. Significant differences were found between different methods for BPF determination. In the MS population, BPF was associated with EDSS. Only progressive disease course could predict EDSS worsening. Conclusion The data on BPF and levels of NFL and GFAP in CSF of healthy individuals can aid in the interpretation of these variables in the setting of MS. Knowledge on differences in BPF data from different methods for BPF determination can be useful in comparing data across studies, but also highlights the need for a commonly accepted gold standard. The correlation between GFAP and NFL levels in CSF may indicate an association between glial and axonal turnover that is independent of the aging effect on the brain. However, the low number of volunteers for LP precluded clear conclusions. An association between BPF and EDSS was seen in the MS group. The ability to predict EDSS worsening in the clinical MS cohort was limited.
Original papers


<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARR</td>
<td>Annualised relapse rate</td>
</tr>
<tr>
<td>$B_0$</td>
<td>The strong magnetic field produced by an MRI scanner</td>
</tr>
<tr>
<td>BPF</td>
<td>Brain parenchymal fraction</td>
</tr>
<tr>
<td>BPV</td>
<td>Brain parenchymal volume</td>
</tr>
<tr>
<td>BSC</td>
<td>Volume within the brain surface contour</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>DIS</td>
<td>Dissemination in space</td>
</tr>
<tr>
<td>DIT</td>
<td>Dissemination in time</td>
</tr>
<tr>
<td>DMT</td>
<td>Disease modifying treatment</td>
</tr>
<tr>
<td>EDSS</td>
<td>Expanded disability status scale</td>
</tr>
<tr>
<td>FS</td>
<td>Functional system</td>
</tr>
<tr>
<td>ELISA</td>
<td>Enzyme-linked immunosorbent assay</td>
</tr>
<tr>
<td>FLAIR</td>
<td>Fluid attenuated inversion recovery</td>
</tr>
<tr>
<td>GBCA</td>
<td>Gadolinium based contrast agent</td>
</tr>
<tr>
<td>GBCA-L</td>
<td>GBCA-enhancing lesions</td>
</tr>
<tr>
<td>GFAP</td>
<td>Glial fibrillary acidic protein</td>
</tr>
<tr>
<td>GM</td>
<td>Gray matter</td>
</tr>
<tr>
<td>ICV</td>
<td>Intracranial volume</td>
</tr>
<tr>
<td>IgG</td>
<td>Immunoglobulin G</td>
</tr>
<tr>
<td>IQR</td>
<td>Interquartile range</td>
</tr>
<tr>
<td>LP</td>
<td>Lumbar puncture</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Meaning</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>MS</td>
<td>Multiple sclerosis</td>
</tr>
<tr>
<td>NEFL, NFL</td>
<td>Neurofilament light</td>
</tr>
<tr>
<td>OCB</td>
<td>Oligoclonal band</td>
</tr>
<tr>
<td>PASAT</td>
<td>Paced auditory serial addition test</td>
</tr>
<tr>
<td>PBVC</td>
<td>Percent brain volume change</td>
</tr>
<tr>
<td>PD</td>
<td>Proton density</td>
</tr>
<tr>
<td>PPMS</td>
<td>Primary progressive multiple sclerosis</td>
</tr>
<tr>
<td>PRMS</td>
<td>Progressive relapsing multiple sclerosis</td>
</tr>
<tr>
<td>RRMS</td>
<td>Relapsing remitting multiple sclerosis</td>
</tr>
<tr>
<td>QoL</td>
<td>Quality of life</td>
</tr>
<tr>
<td>SIENAX</td>
<td>Structural Image Evaluation Using Normalisation of Atrophy Cross-Sectional</td>
</tr>
<tr>
<td>SyMap</td>
<td>Synthetic tissue mapping</td>
</tr>
<tr>
<td>SPM</td>
<td>Statistical parametric mapping</td>
</tr>
<tr>
<td>SPMS</td>
<td>Secondary progressive multiple sclerosis</td>
</tr>
<tr>
<td>VBM</td>
<td>Voxel based morphometry</td>
</tr>
<tr>
<td>WM</td>
<td>White matter</td>
</tr>
</tbody>
</table>
Enkel sammanfattning på svenska

Multipel Skleros (MS) är en neurologisk sjukdom som beror på att det egna immunsystemet angriper hjärnan och ryggmärgen. Detta kan leda till neurologisk funktionsnedsättning och vid magnetkameraundersökning (MR-undersökning) av hjärnan och ryggmärgen kan inflammatoriska skador och förlust av nervvävnad ses. Förlust av nervvävnad i hjärnan betecknas ofta hjärnatrofi. Precis som många andra organ i kroppen så förändras hjärnan genom livet, och hjärnatrofi sker genom det naturliga åldrandet även hos friska personer. Hos individer med MS sker detta dock, på grund av sjukdomen, i högre takt än normalt. MS-orsakad hjärnatrofi har i ett flertal studier visat sig ha en betydelse för hur MS-sjukdomen utvecklas och kan dessutom påverkas av MS-specifik behandling. Av denna anledning introducerades uppföljning av hjärnatrofi, mätt med MR, i vårdprogrammet för individer med MS vid Norrlands Universitetssjukhus 2009/2010.

För att kunna tolka mått på hjärnatrofi vid MS är det viktigt att veta normalvärdet hos friska personer. Denna avhandlings syfte var därför att undersöka mått på hjärnatrofi hos friska för att öka kunskapen om detta, och särskilt i jämförelse mot individer med MS. Ett ytterligare syfte var att undersöka om det finns en koppling mellan hjärnatrofi och mätning av specifika cellskademolekyler i vätskan runt hjärnan och ryggmärgen, den s.k. cerebrospinalvätskan. Ett tredje syfte var att undersöka om det i den kliniska uppföljningen av MS går att förutsäga vilka individer som kommer att försämras i sin sjukdom i framtiden samt om samma prognostiska värde av mått på hjärnatrofi som setts i tidigare studier även ses inom den kliniska uppföljningen.

Avhandlingen visar hur utveckling av hjärnatrofi förändras genom livet, samt att det finns skillnader mellan olika sätt att mäta detta på. Det påvisades en koppling mellan cellskademolekyler i cerebrospinalvätska och mått på hjärnatrofi. Studieresultaten talade dock för att denna koppling berodde på att båda dessa mått är kopplade till personens ålder. Grad av hjärnatrofi var associerat med funktionsnedsättning vid MS, vilket överensstämmer med vad som setts i tidigare vetenskapliga studier. Det var mycket svårt att förutsäga vilka individer med MS som skulle försämmas.
Introduction

Multiple Sclerosis

Epidemiology and clinical course

Multiple sclerosis (MS) is an autoimmune disease of the central nervous system (CSN). Studies conducted at national or regional levels in Sweden have reported an incidence of 6.0 to 10.2 cases per 100,000 persons and a prevalence of 188 to 215 per 100,000 (1-3).

MS is characterised by subacute neurological dysfunction caused by focal inflammatory activity, often called relapses, and/or a progressive accumulation of disability (4). An MS relapse can generally be defined as new or worsened neurological symptoms that persist for at least 24 hours and cannot be better explained by other causes (5). An MS-relapse may regress completely but may also result in permanent disability (6). Progressive accumulation of disability typically follows a pattern of insidious, worsening of symptoms over months to years, leading to increasing neurological deficits that are usually permanent.

The most common presentation of MS at onset is focal inflammatory activity, but without progressive worsening. This is termed Relapsing Remitting MS (RRMS). For reasons that are unclear, many individuals that initially present with relapses, convert to a disease course dominated by progressive worsening later on, denoted as Secondary Progressive MS (SPMS). A less common course is characterised by progressive worsening from onset, without relapses, called Primary Progressive MS (PPMS). The term Progressive Relapsing MS (PRMS) has been used to denote disease presenting with concomitant relapses and progressive worsening.

The terms used to describe the clinical course of MS were redefined recently according to an international consensus (4). The term PRMS was abandoned and instead it was acknowledged that each MS disease course can be described in terms of absence or occurrence of focal inflammatory activity along with absence or occurrence of progressive accumulation of disability (Table I). The term RRMS is still used for MS presenting exclusively with relapses. The terms SPMS and PPMS are used somewhat more liberally to encompass mixes of progressive worsening and relapses.
### Relapsing disease course

<table>
<thead>
<tr>
<th>Clinically isolated syndrome (CIS)</th>
<th>Active</th>
<th>Not active</th>
</tr>
</thead>
<tbody>
<tr>
<td>First symptoms raising the suspicion of MS, but without fulfilling diagnostic criteria</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relapsing Remitting MS (RRMS)</th>
<th>Active</th>
<th>Not active</th>
</tr>
</thead>
<tbody>
<tr>
<td>The most common disease course at onset, characterised by intermittent inflammatory activity, clinically identified as relapses</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Progressive disease course

<table>
<thead>
<tr>
<th>Either</th>
<th>Active with ongoing progression</th>
<th>Not active but ongoing progression</th>
<th>Active but without ongoing progression</th>
<th>Neither active nor ongoing progression (stable disease)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primary Progressive (PPMS)</strong></td>
<td>(Progressive accumulation of disability from onset)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Secondary progressive (SPMS)</strong></td>
<td>(Progressive accumulation of disability after an initial debut of RRMS)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 1: MS disease courses

The table presents disease courses of MS as defined by the consensus statement (4). “Active” MS is defined as MS with occurrence of clinical relapses and/or MRI activity. “Progression” is defined as clinical evidence of increased neurological impairment independent of relapses. Assessments of activity/progression should be performed at least annually. If assessment is not available the activity/progression is defined as “indeterminate.”
The use of MRI in MS

By recording how hydrogen atoms within tissues of the body respond to magnetic fields applied outside of the body, the tissue in question can be visualised as a grayscale image via the process of magnetic resonance imaging (MRI). MRI is a non-invasive imaging technique that can provide detailed images of the CNS. Image resolution and tissue contrast are influenced by magnetic field strength as well as how the hydrogen atoms are manipulated. While the magnetic field strength of a particular MRI-scanner cannot be changed, the magnetic manipulation of the hydrogen atoms can be adjusted and used to obtain different types of MR-image, such as T₁ weighted, T₂ weighted and Fluid attenuated inversion recovery (FLAIR) images, which are of special interest in MS (see the section on MRI for further details, pages 7-12).

MRI can be used to visualise focal inflammatory activity, or ‘lesions’, in MS. These may be localised in brain or spinal cord white matter (WM) or gray matter (GM). WM lesions are most easily visualised, presenting as bright, or hyperintense, areas on T₂ weighted images. It is more difficult to detect GM-lesions, although it has been shown to be easier with higher MRI-scanner field strengths (7). MRI lesions can occur independently of relapses (8). The ratio of MRI lesions to clinical relapses can from MS clinical trial data be approximated to about 4-12 lesions per relapse (9-21), although the variation reported between studies is large. MS may also present itself initially by typical MRI findings without any clinical symptoms (22).

Another tool in the MRI-follow-up of MS is administration of a Gadolinium-based contrast agent (GBCA) (23). Recent or ongoing inflammatory activity may be associated with disruption of the blood-brain-barrier, allowing GBCA to leak out into the brain parenchyma. This can be visualised on T₁ weighted images as hyperintense (bright) areas (23).

Analysis of CSF in MS

The analysis of CSF is widely used in the investigation of suspected MS. While not required for the diagnosis of RRMS or SPMS, it can support the diagnosis of PPMS (24). Furthermore, it can help in the assessment of differential diagnoses, such as CNS infection. The CSF investigations of most importance are the Immunoglobulin G (IgG) index, defined as the ratio of the IgG concentrations in CSF and in serum divided by the ratio of albumin concentrations in CSF and in serum (25), and the presence or absence of oligoclonal bands (OCBs). The presence of OCBs indicates the existence of
IgG producing plasma cell clones in the CSF that are not present in serum (25). Analyses of other markers of inflammation or tissue damage in the CSF can also be useful in MS. The tissue damage-related markers neurofilament light polypeptide (NEFL; commonly referred to as NFL) and glial fibrillary acidic protein (GFAP) will be discussed in a later section (pages 19-20).

The diagnosis of MS

No clinical, radiological or laboratory finding can be used to diagnose MS unequivocally. Instead, the diagnostic process relies on various parameters including clinical symptoms and radiological findings. Current MS diagnostic criteria, referred to as the McDonald criteria, were most recently revised in 2010 (24) (Table II). The McDonald criteria reflect the idea that MS is a chronic disease that can potentially affect any anatomical area of the CNS. Therefore, criteria are based on the demonstration of the two criteria of dissemination in time (DIT), i.e. evidence of MS disease activity at more than one point in time, and dissemination in space (DIS), i.e. evidence of MS disease activity in at least two distinct areas of the CNS. Following the occurrence of at least one episode of clinical symptoms and signs indicative of MS, a diagnosis can be made if both DIS and DIT can be demonstrated. SPMS can be diagnosed in the setting of progressive accumulation of disability in a patient initially debuting as RRMS.

Diagnosis of PPMS requires special consideration. For PPMS, DIT has been replaced by at least one year of progressive clinical worsening. An elevated IgG index, or presence of OCBs, can support the diagnosis of PPMS.
### Relapsing Remitting MS (RRMS)

<table>
<thead>
<tr>
<th><strong>MRI</strong></th>
<th><strong>Clinically</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dissemination in space</strong></td>
<td>≥1 T2-lesion in at least 2 out of the 4 areas:</td>
</tr>
<tr>
<td></td>
<td>- Periventricular</td>
</tr>
<tr>
<td></td>
<td>- Juxtacortical</td>
</tr>
<tr>
<td></td>
<td>- Infratentorial*</td>
</tr>
<tr>
<td></td>
<td>- Spinal cord*</td>
</tr>
<tr>
<td><em>In the case of brainstem or spinal cord syndrome the symptomatic lesion is not counted.</em></td>
<td>≥2 clinical relapses from at least 2 different anatomical locations.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Dissemination in time</strong></th>
<th>Either:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≥1 new T2- and/or GBCA-enhancing lesion at follow-up MRI after baseline MRI where at least one MS-suspected lesion has been identified.</td>
</tr>
<tr>
<td>Or:</td>
<td>A concomitant finding of asymptomatic GBCA-enhancing and non-GBCA-enhancing lesions.</td>
</tr>
<tr>
<td></td>
<td>≥2 clinical relapses at different time points.</td>
</tr>
</tbody>
</table>

### Primary progressive MS (PPMS)

1. One year of progressive worsening
And:
2. At least 2 out of the following:
   - Evidence for dissemination in space based on ≥1 T2-lesion in at least 2 localisations typical for MS (periventricular, juxtacortical, infratentorial).
   - Evidence for dissemination in space based on ≥2 T2-lesions in the spinal cord (symptomatic lesions disregarded).
   - Findings of oligoclonal bands or elevated IgG-index.

### Table II: The McDonald criteria (2010 revision)

The McDonald criteria (24) (2010 revision) for diagnosing RRMS and PPMS. The table is translated and adapted from the national MS-MRI guidelines from the Swedish MS association (26). The diagnosis of Secondary progressive MS is based on the occurrence of progressive worsening after an initial diagnosis of RRMS.

*GBCA = Gadolinium based contrast agent*
Expanded Disability Status Scale (EDSS)

Neurological impairment in MS can affect any part of the CNS. A widely used method to quantify functional impairment is via a standardised neurological examination and scoring according to the Expanded Disability Status Scale (EDSS) (27). This is used to calculate a score for each of eight predefined functional systems (FSs) attributable to MS; visual, brainstem, cerebellar, pyramidal, sensory, bowel and bladder, cerebral and other functional impairment. A higher FS score corresponds to higher degree of functional impairment of the specific FS. These are then combined with the reported or measured maximum walking distance to produce a compound EDSS score ranging from zero (no disability) to ten (death caused by MS).

Advantages of the EDSS include the relative ease of scoring for an experienced user of the scale and the fact that no special equipment is needed beyond what is usually available for a standard neurological examination. However, despite its advantages and widespread use, the scale is not free from critique. Points of criticism include the fact that even though a higher EDSS score generally implies a larger degree of functional impairment; the scale is ordinal, making statistical calculations based on means or medians difficult to interpret. The scale is also prone to measurement variations (28-30). Another factor to keep in mind is that walking impairment will have a relatively large impact on the EDSS score compared with other symptoms. Fatigue and impairment of cognitive functions are notable examples, as they contribute relatively little to the compound EDSS score while they may independently impact a patient’s quality of life (31-33).

Clinical follow-up of MS

The most common way of assessing disease activity and current prognosis are regular standard neurological examinations with the addition of repeated MRI examinations (23). A primary aim of MRI examinations is to identify signs of new focal inflammatory lesions, detectable as hyperintense lesions on T₂ weighted images, and ongoing inflammation, detectable as hyperintense areas on T₁ weighted images after administration of GBCA.
Magnetic Resonance Imaging

The concept of MRI, explained briefly

The basic idea of medical MRI is that hydrogen atoms, abundant in biological tissues, can act as magnetic dipoles (34). As such they can be manipulated by use of external magnetic fields. Their own magnetic behaviour, following such a manipulation, depends on both the local density of hydrogen atoms and their molecular surroundings. As the magnetic behaviour of hydrogen atoms can be measured, biological assumptions regarding the site of manipulation and recording can be made non-invasively. This makes MRI a powerful medical imaging tool, which is used extensively for imaging of the CNS. What will be presented here is only a simplified theoretical model to assist in the description of basic MRI methodology.

The hydrogen atom and the MRI scanner

The following basic description of MRI (pages 7-11) is based on the book ‘MRI: From Picture to Proton, 2nd Ed.’ by McRobbie et.al. (35).

Two important inherent properties of the hydrogen atom are that its nucleus consists of a single positively charged proton and that it continually rotates around its own axis, the latter being referred to as atomic spin. This means that the hydrogen nucleus constitutes a moving electrical charge and is associated with its own, small, electromagnetic field, i.e. it is a magnetic dipole. The magnetic field of such a dipole can be viewed as a vector, i.e. it has a direction and a magnitude. In the presence of a strong external magnetic field, the direction of the magnetic vectors of hydrogen nuclei will align themselves to its direction. This is what happens when hydrogen nuclei are placed in an MRI-scanner (Figure I). A theoretical model for this alignment is that of parallel versus anti-parallel orientations. This model suggests that the magnetic vectors of all hydrogen nuclei affected by the strong magnetic field of the MRI scanner (called B₀) will align along the same axis as B₀ but can be oriented either in the same direction (parallel) or the opposite direction (anti-parallel). A larger proportion of nuclei will align themselves parallel to B₀. Therefore, the sum of all magnetic vectors of hydrogen nuclei will amount to a net magnetic vector in the same direction as B₀ (Figure II).
Figure I: Hydrogen atoms at rest
At resting state the hydrogen nuclei (blue spheres) and their corresponding magnetic vectors (red arrows) are not giving rise to any meaningful net magnetic vector.

Figure II: Hydrogen atoms in an external magnetic field
When exposed to a strong magnetic external field (the bold red lines at the right and left side of the image) such as that in an MRI scanner, the individual magnetic vectors of the hydrogen nuclei will align parallel or anti-parallel to the external field. As a larger proportion will align parallel to the external field, a net magnetic vector is produced by the sum of all the small magnetic vectors.
Another important concept for understanding how MRI works is that of precession. This refers to the fact that each hydrogen nucleus, as aligned to $B_0$, is not perfectly aligned to the exact direction of $B_0$ but rather rotates around the axis of $B_0$ at a slight tilt, comparable to the motion of a spinning top (Figure III).

**Precession**

The small net magnetic vector that arises from the sum of all the hydrogen nuclei when subjected to $B_0$ contains information about the tissue. The field strength of $B_0$, temperature and the local density of hydrogen nuclei influence the magnitude of this net vector, referred to as “longitudinal magnetisation”. However, as this is parallel to $B_0$, which due to its much higher relative field strength overshadows the small net magnetisation from hydrogen nuclei, it is difficult to measure. In order to facilitate recording of the net magnetic vector of hydrogen nuclei they can be manipulated by applying a second, temporary external magnetic field (called a radiofrequency pulse), perpendicular to $B_0$. This has two important effects; it flips the net magnetic vector of hydrogen nuclei out of alignment with $B_0$ and forces them to synchronise their precessional movements. This synchronised precession means that the magnetic vectors of all hydrogen nuclei will share a common direction even when forced out of alignment with $B_0$, resulting in

**Figure III: Precession**

*The figure shows a spinning top (gray conical volume), rotating around its own axis (solid black line) and at the same time rotating, or precessing, around an imaginative longitudinal axis (central dashed line). The same concept applies to a hydrogen atom that has aligned itself with a strong external magnetic field.*
Figure IV; Longitudinal and transversal magnetisation

When a large amount of hydrogen nuclei (blue sphere), each with a precessing magnetic vector (red diagonal arrow), are placed in a strong magnetic field ($B_0$) a net magnetic vector (red vertical arrow) arises. This is called the longitudinal magnetisation. When the perpendicular magnetic field, called the radio frequency pulse (not shown), is applied the atoms are forced out of alignment with $B_0$ and the longitudinal magnetisation is diminished. At the same time, the radio-frequency pulse forces the nuclei to synchronise their precessions, creating the transversal magnetisation (red horizontal arrow). When the radiofrequency pulse is terminated these processes are reversed; the system ‘relaxes’.

a net magnetic vector that is not aligned with $B_0$ and thus can be recorded. This magnetic vector is referred to as “transversal magnetisation”. At the same time, the component of longitudinal magnetisation will be diminished as a result of the radiofrequency pulse, as the hydrogen nuclei are no longer aligned to $B_0$ (Figure IV). This is important for creating the MRI-images, which will be described further below.

Relaxation

Upon cessation of a radiofrequency pulse the net magnetic vector of hydrogen nuclei will be out of alignment with $B_0$ and will display a synchronised precession. However, released from the effects of the radiofrequency pulse, they will begin to realign to $B_0$ and precessional synchronisation will start to decay. After a short period, the net magnetic vector of hydrogen nuclei will have realigned fully to $B_0$ and precessional movements will have returned to being fully asynchronous. This means that
the longitudinal magnetisation will now have been reconstituted and the transversal magnetisation will now have been lost. This process is called ‘relaxation’ and can be measured.

By measuring the relaxation, several important MRI tissue parameters can be extracted. The three main physical attributes of a tissue that are measured are:

- **T₁ relaxation time**
  The time required for 63% of the longitudinal magnetisation to be reconstituted after the radiofrequency pulse is stopped

- **T₂ relaxation time**
  The time required for 63% of the transversal magnetisation to be lost after the radiofrequency pulse is stopped

- **Proton density (PD)**
  The magnitude of the longitudinal magnetisation in the relaxed state

The density of hydrogen nuclei and the molecular constituents of the tissue in question affect these three parameters differently. By introducing a sequence of radio-frequency pulses of different periods and carefully choosing the timing of recording, a signal can be recorded that is more or less influenced by one of these three parameters.

If a volume of biological tissue is overlaid with a 3D grid comprised of small cubes or rectangles (termed voxels) and the same recording is made for each of these voxels, they can be combined to create a 3D image that is:

- **T₁ weighted**
  The contrast between different voxels is heavily influenced by the difference in T₁ relaxation time between the voxels.

- **T₂ weighted**
  The contrast between different voxels is heavily influenced by the difference in T₂ relaxation time between the voxels.

- **PD weighted**
  The contrast between different voxels is heavily influenced by the difference in PD between the voxels.

While complex sequences of radiofrequency pulses can provide a variety of different image types and weightings, the three types of MR images mentioned above are the primary elements of traditional MR imaging. Apart from these, a widely used image type in the setting of MS care and research is the FLAIR image. The tissue contrast of a FLAIR image is similar to that of a T₂ weighted image, with the important difference that the tissue signal
from water (i.e. CSF) has been voided by use of a specific radiofrequency pulse sequence. This means that CSF, while otherwise very bright on T₂ weighted images, will appear dark. This increases the ability to discern the lesions, which are also bright on T₂ weighted images, if these are located close to areas with CSF. As periventricular lesions are common in MS, this is a useful image type.

**Limitations of conventional MRI**

Although MRI is a very useful technique, the conventional application of MRI, as outlined above, is associated with some limitations. One example is that the contrast in conventional MR imaging is based on the relative difference in, for example, T₂ relaxation time between voxels but the absolute values are not determined. This means that more information about the examined tissue could potentially be gained if absolute quantification of tissue parameters would be performed. Such quantification have previously been held back by long scan times, but technical advancements have now reduced the time requirement considerably (36).

Another limitation of conventional imaging is the reliance on a specific predefined MRI protocol for each differently weighted image that is to be acquired. If both a T₁ and a T₂ weighted image are required, two different MRI protocols need to be run. If multiple different image types are needed, as may be the case for a clinical investigation, the total scan time increases. This makes MRI a time consuming imaging method. This also means that if a conventional MRI examination would need to be supplemented with another image type after the original examination, the person examined would need to be rescanned.

**Synthetic Tissue Mapping**

Synthetic Tissue Mapping (SyMap) is a different approach to MRI that aims to overcome both of the limitations of conventional imaging described above. It is based on a specialised MRI acquisition protocol (37) that allows for the determination of the absolute values of T₁ and T₂ relaxation times as well as PD, with a relatively short scan time (approximately 6 minutes) for each voxel in the image (38). This information can be analysed using dedicated computer software (SyMRI Brain Studio, SyntheticMR AB, Linköping, Sweden) to automatically define and segment the different tissue types of the image. By the use of PD and the relaxation rates R₁ and R₂, defined as the inverse of the T₁ and T₂ relaxation times, respectively, a 3D space with R₁, R₂ and PD as axes can be established. In this space, the position of normal GM, WM and CSF have been defined (39). In an examination of the brain, the
software assigns each voxel in the image a proportionate amount of GM, WM, CSF and NON (tissue not similar enough to any of the three other classes to be designated as such) based on its position in $R_1$-$R_2$-PD space (Figure V). This allows for automatic quantification of the volumes of these predefined tissue types. This information can be utilised in assessing the occurrence of brain atrophy (see below). It can also be used to create synthetic versions of conventional MR image types, meaning that multiple different MR image types, e.g. $T_1$, $T_2$, PD weighted, can be created as needed by post processing of a single MRI acquisition (38).

Figure V: Synthetic Tissue Mapping (SyMap)
The major tissue types in the brain; gray matter (green), white matter (blue) and cerebrospinal fluid (purple), differs in tissue composition and therefore also in the properties measured by MRI. Based on the differences in $T_1$ relaxation, $T_2$ relaxation and proton density each individual voxel can be classified as belonging to either or several of the three tissue types. The above graph exemplifies the tissue locales in $R_1$ and $R_2$ space. $R_1$ and $R_2$ are defined as $1/T_1$–relaxation time and $1/T_2$–relaxation time respectively.
Brain atrophy

Most of the organs and tissues of the body undergo changes associated with the normal aging process. The brain is no exception and it has been widely reported that normal aging is associated with a decrease in the volume of brain tissue (40-47). The annual percentage of brain volume change (PBVC) (loss of brain volume) is approximately 0.2% to 0.3% per year in healthy individuals of 21 to 60 years of age (48, 49), although rates of up to 0.66% per year have been reported in a cohort of individuals from 55 to 90 years old (50). The cause of this age-related atrophy is largely unknown.

The loss of brain parenchymal volume (BPV) at a higher rate than that seen in normal aging, i.e. pathological brain atrophy, is an important feature in several neurodegenerative disorders such as dementia (51), Huntington’s disease (52) and MS (53). However, the loss of brain parenchyma with normal aging has not been studied in sufficient detail to draw clear conclusions regarding how pathological brain atrophy might differ from the loss of brain parenchyma that is seen in normal aging.

**Brain atrophy in MS**

Brain atrophy can be detected in all stages and types of MS and an association between pathological brain atrophy and MS was described at least as early as 1938 (54). As more sensitive methods to study brain atrophy have been developed, it has generated increasing interest (54). A paper published 1999 by Rudick et.al further spurred this interest after reporting that brain atrophy in MS could be attenuated by use of a disease modifying treatment (DMT) (53). It is known that the degree of whole brain atrophy in MS is associated with functional impairment (53, 55-58) and it has been further established that the rate of atrophy can be reduced by use of DMTs (59-62). A meta-analysis on the effect of DMTs on brain atrophy has reported an annual PBVC in treated study arms of 0.27% to 0.33% (divided based on choice of DMT to compare one high efficacy and one low efficacy DMT-group)(62). The PBVC in the placebo arms was estimated to 0.50% per year (62). Higher annual PBVC have also been reported (48, 62, 63), up to 1.5% and 2.0% per year in untreated RRMS and SPMS populations respectively (63).

Assessment of brain atrophy may be associated with MS disability independently from quantitative measurements of inflammatory lesions (54, 64). It is generally included as an MRI outcome in pharmaceutical trials of MS treatments (12, 17, 20, 21, 61, 62). Perhaps even more importantly, it has been shown that rate of brain atrophy during longitudinal observation is
associated with functional outcome (49, 55, 65-70) and conversion from first presenting symptoms to clinically definite MS fulfilling diagnostic criteria (69, 70). Furthermore, the treatment effect seen on brain atrophy with DMTs correlates with the treatment effect on disability (59) and has been reported as independent of the effect of DMTs on inflammatory lesions and clinical relapses (60). This implies a clinical value of information from brain atrophy measurements in MS that is independent from what is gained from assessment of other MRI-biomarkers. Efforts have been made to define clinical cut-off values for atrophy rate to distinguish pathological from healthy rates and to predict clinical worsening, indicating, that an annual PBVC of 0.52% per year may distinguish pathological atrophy (49).

While whole brain atrophy is the aspect of atrophy most commonly included in clinical trials and perhaps the one most thoroughly studied in MS, regional brain atrophy as well as spinal cord atrophy has also been investigated and reported to be associated with several different aspects of functional impairment attributable to MS (64, 71).

**The cause of MS-associated brain atrophy**

The underlying causes of brain atrophy in MS are not well understood (49, 54, 72). Correlations between brain atrophy and inflammatory lesion activity have been reported (58, 73, 74). Therefore, it is plausible that demyelination and axonal damage associated directly with the formation of lesions play a role in the development of brain atrophy (54). However, other pathological processes separate from focal inflammatory activity cannot be ruled out, especially since brain atrophy has been associated with disability independent of inflammatory lesions and clinical relapses (59, 60). Furthermore, correlations between inflammatory activity and assessments of brain atrophy are only moderate in strength (54, 55, 58). More prominent brain atrophy has also been reported in PPMS compared to RRMS, despite the disease duration being similar between groups (75). As PPMS is dominated by progressive accumulation of disability, while RRMS is characterised by focal inflammation, this further supports the notion that MS-associated brain atrophy and inflammatory lesions may, at least in part, represent different pathophysiological aspects of the disease. However, a caveat in this observed discrepancy between inflammation and neurodegeneration is that commonly used MRI field strengths may not be sufficient to visualise the entirety of the inflammatory lesion load (7).
Quantification of brain atrophy

Various methods are available to quantify brain atrophy (61, 76). All are to some extent relying on distinguishing, or segmenting, the MRI brain data into different tissue types, as was exemplified with SyMap in the section on MRI above. One of the simplest methods to quantify atrophy is thus to directly determine the volume of BPV. However, an issue in quantifying differences in BPV is that there are large inter-individual variations not related to pathology (40). These variations can be decreased if the BPV is normalised to the volume of the intracranial cavity (ICV) (40, 76). The ICV varies only minimally during the adult life (77, 78) and can thus provide a robust index for normalisation, making meaningful interpretations regarding brain atrophy easier.

Brain Parenchymal Fraction (BPF)

One way of normalising BPV to the ICV is by calculating the ratio of BPV divided by ICV. This ratio was previously referred to as the percentage of brain parenchyma (79). The term brain parenchymal fraction (BPF) was later introduced as the ratio of BPV to volume within the brain surface contour (BSC) (53). The delineation of the surface contour of the brain to produce the volume to be used as denominator (53) does by definition not yield the exact same ratio as BPV to ICV (Table III, Figure VI). Despite this, the term BPF have transitioned into most commonly referring to the ratio of BPV to ICV and the term percentage of brain parenchyma is not generally used. However, it must be emphasised that no clearly defined consensus exists regarding this nomenclature. BPF is sometimes used according to the original definition and the ratio of BPV to ICV may sometimes be referred to by other terms than BPF.
**Original definition of BPF**

\[
BPF = \frac{\text{Brain parenchymal volume}}{\text{Volume within the brain surface contour}}
\]

Most widely used definition of BPF, originally referred to as percentage of brain parenchyma

\[
BPF = \frac{\text{Brain parenchymal volume}}{\text{Total intracranial volume}}
\]

**Table III: Definition of brain parenchymal fraction (BPF)**

The term BPF was originally used for the ratio between brain parenchymal volume (BPV) and the volume that was contained within a delineation of the brain surface contour in a 2D MR-image. The term BPF has transitioned into most often referring to the ratio between BPV and total intracranial volume.

**Figure VI: BPF according to the original and the now most widely used definition**

The original definition of BPF (left image) uses the delineation of the outer surface of the brain (red line) as the denominator for the ratio. The now most widely used definition (right image) instead uses the delineation of the total intracranial space. \( BPF = \text{Brain parenchymal fraction} \)
**Issues associated with the quantification of brain atrophy**

Quantification of brain atrophy involves problems associated with methodological variations as well as physiological factors influencing the measurements. Technical variability varies with choice of method and may also vary with MRI scanner. One study has reported intra- and inter-scanner coefficients of variability for brain volume measurements using the methods Structural Image Evaluation Using Normalisation of Atrophy Cross-Sectional (SIENAX), Statistical Parametric Mapping (SPM) and Voxel Based Morphometry (VBM) to range between 0.38% and 1.42% (80). For the SyMap method the user manual supplied by the manufacturer reports the scan to rescan variability of BPF to be between 0.3% and 1.3% for different MRI scanners, as investigated by internal testing on 12 healthy subjects (81). However, a study investigating a larger population of healthy individuals (n=40) as well as individuals with MS (n=20) reported intra-scanner coefficients of variation of 0.23% and 0.25% for the two groups, respectively (82). This methodological variation is not negligible if related to the reported values for annual PBVC in MS and healthy individuals mentioned above.

Hydration status may affect brain volume measurements before and after 16h fasting (83, 84). However, another study that tried to replicate a clinically probable scenario with overnight fasting (9h) could not detect any significant differences (85). Furthermore, it has been reported that time of day may be of importance in brain volume measurements (86), albeit with a rather small effect size that could not be reproduced in a smaller study (n=2) (80).

When specifically considering MS, an increased rate of brain atrophy has been reported directly following initiation of disease modifying treatment (87, 88). This increased rate has been correlated to the level of baseline inflammatory activity (88), supporting the idea of pseudoatrophy, i.e. a quick reduction in brain volume after initiation of therapy caused by decreased inflammation-associated oedema (87).

Treatment of an MS relapse with intravenous methylprednisolone may influence measurement of brain volume shortly after treatment (89). One study has reported a significantly reduced brain volume up to two months after treatment (90). Another could not detect any significant effect as close as 20 days after treatment, but did see a significant effect after 30 days in a subgroup of patients which had received oral steroid tapering after the initial infusion (91). The background for this transient effect on brain volume is not fully understood, but changes to vascular permeability and decreased water content of the brain may be part of the explanation (91).
All these factors influence the assessment of brain atrophy, adding to the complexity of the measurement and its interpretation.

**CSF analyses in MS**

Apart from the identification of an elevated IgG index and detection of OCBs (25), a wide range of CSF biomarkers have been studied in MS (92). Two markers of interest are NFL and GFAP.

*Neurofilament light (NFL)*

Neurofilaments are heteropolymers inherent to the structure of axons, both in the peripheral and central nervous systems. They consist of neurofilament light (NFL), medium and heavy chain polypeptide subunits (93). Axonal damage in the CNS leads to an increase in the concentration of NFL in the CSF (94-96), and elevated levels can be measured using an enzyme-linked immunosorbent assay (ELISA)(97).

The levels of NFL in the CSF are increased in MS and are particularly high during concurrent disease activity, such as clinical relapses (98-100). Published values of NFL in healthy individuals include 350 ng/l in a young (98) and approximately 1150 ng/l in an elderly population (101). There is a correlation between level of NFL in CSF and age in healthy individuals (101-105).

*Glial fibrillary acidic protein (GFAP)*

Glial fibrillary acidic protein (GFAP) is a component of the astrocyte cytoskeleton. In the event of astrocyte damage GFAP may be released into the CSF and elevated levels of GFAP can be identified using ELISA (96).

In MS, elevated levels of GFAP are primarily seen in the progressive forms of the disease (92, 98, 99). However, elevated levels have been reported in RRMS after a relapse (99). A mean level of GFAP in healthy, young individuals of 435 ng/l has been reported (99). Correlation between GFAP and age has been reported (106)
The relationships of NFL and GFAP to age

The levels of both NFL and GFAP in CSF have been reported to correlate with age (101-106). However, the data in healthy individuals is limited. In addition, since brain atrophy determined by MRI also correlates with age (40-47), it could be hypothesised that the reasons for the correlation of each of these biomarkers with age represent the same biological effect of ageing on CNS. The relationship between CSF and MRI markers of neurodegeneration is insufficiently studied.

The clinical care program for MS at Umeå University Hospital, Sweden.

Umeå University Hospital is a regional hospital located in Umeå, Västerbotten County, Sweden. It provides specialised hospital care for the residents of the northern part of Sweden. The neurology clinic at Umeå University Hospital functions as a primary MS care centre for all incident MS cases in Västerbotten and as a secondary and tertiary MS care centre for the remainder of Northern Sweden.

The clinical care program for MS patients at Umeå University Hospital includes regular neurological examinations, during which the determination of EDSS is generally performed. Regular MRIs are performed for all patients with an active disease and/or on treatment with DMTs. In 2009/2010 the clinical care program for MS was revised, and assessment of BPF using SyMap was added to the follow-up of MS at each MRI examination.

Rationale of this dissertation

BPF and levels of NFL and GFAP in CSF in healthy individuals

BPF, as well as NFL and GFAP levels in CSF are all indicators of CNS tissue damage with an association to the person’s age. In order to properly interpret values of these variables in the setting of disease, the expected impact of the normal aging process must be known.

Comparing different methods for BPF determination

There is no gold standard for BPF determination, but several methods exist. All generally rely on segmenting an MRI examination of the brain into different tissue types. This can be done either fully manually or by using a computer software to perform parts of or the whole segmentation. Fully manual image segmentation has the advantage of providing more control
over the segmentation process but is time-consuming. Furthermore, manual image segmentation carries a risk for operator-dependent bias while automated and semi-automated methods are prone to method-dependent biases. In order to properly interpret results from studies that have employed different segmentation methods, they first need to be compared in relation to one another.

Can the prognostic value of BPF be reproduced in clinical follow-up?

The clinical care program for MS in Umeå includes follow-up of BPF using SyMap. However, despite the biological importance of brain atrophy reported in clinical trials (49, 55, 65-70), the application of clinical follow-up of atrophy in MS must be evaluated systematically in order to ensure that the associations with disability and prognosis can be replicated in a real life clinical setting.

Aims of this dissertation

This dissertation had several specific aims:

- To investigate the normal values of BPF and levels of NFL and GFAP in the CSF and their respective relationship with age
- To investigate the relationships between BPF and levels of NFL and GFAP in the CSF
- To investigate the differences in BPF-values obtained with different MRI post processing methods of common use in MS research
- To use observational data from the clinical care program at Umeå University hospital to ascertain if the prognostic value of brain atrophy follow-up in MS, which have been reported in previous studies, can be reproduced
Materials and methods

Ethical statement

The work presented in papers I, II (Dnr: 2011-39-31M and amendments) and IV (Dnr: 2014-43-31 and amendments) was reviewed and approved by the regional Ethical Review Board of Umeå University. All the healthy volunteers for MRI and/or lumbar puncture provided written informed consent for participation.

The investigation of the MS population in Paper IV used only retrospective data collected through the process of routine clinical follow-up, for anonymised presentation. Therefore, informed consent was not considered necessary, by decision from the regional Ethical Review Board. The clinical care program has been reviewed and approved by the regional ethical review board (Dnr: 2010-351-31).

Paper III only encompassed a review of the literature and was thus not subjected to review by the regional ethical review board.
Study populations

Healthy volunteers
(Papers I, II and IV)

Volunteers without a diagnosis of a neurological disease and without first-degree relative with such a disease were recruited. This was performed via both local advertisements at the neurology clinic at Umeå University Hospital and via newspaper advertisements. A qualified research nurse (Erika Figaro) interviewed all prospective participants in order to assess inclusion and exclusion criteria. The research nurse consulted one of the study physicians (Anders Svenningsson or Mattias Vågberg) as needed. For all participants that volunteered for an MRI examination, the examination result was reviewed by a senior consultant in neuroradiology (with rare exceptions either Richard Birgander or Thomas Lindqvist) and a senior consultant in neurology (Anders Svenningsson).

The criteria for inclusion (for lumbar puncture and/or MRI) were:
1. Ability to participate and provide informed consent.
2. Age above 18 years old.

The criteria for exclusion (for lumbar puncture and/or MRI) were:
1. Symptoms, findings or history indicative of disease suspected to cause brain tissue injury or first degree relatives with such a disease.
2. Contraindications for undergoing lumbar puncture and/or MRI.

A total of 106 healthy individuals (56 females, 50 males, median age 45.7 years, IQR 28.3) were included and underwent an MRI examination. A total of 53 individuals (27 females, 26 males, median age 33.7 years, IQR 21.3) underwent lumbar puncture. A subset of 48 individuals (24 females, 24 males, median age 33.0 years, IQR 18.4) volunteered for both MRI and lumbar puncture, thus, these represent an overlap between the two populations (Figure VII).

In Paper II, a subset of 54 individuals (34 females, 20 males, median age 49.2 years, IQR 33) was selected randomly for manual BPF estimation from the 106 individuals having undergone MRI.
Figure VII: The healthy volunteers for MRI and lumbar puncture
A total of 111 healthy individuals volunteered to participate in the studies. A total of 106 volunteered for MRI, 53 for lumbar puncture and a subset of 48 individuals volunteered for both MRI examination and lumbar puncture.

MRI = Magnetic resonance imaging
**Individuals with MS**  
*(Paper IV)*

The majority of patients receiving care for MS at Umeå University Hospital are included in the Swedish MS Registry (www.neuroreg.se), a national patient registry. The Swedish MS Registry contains information on diagnosis, EDSS scores, relapses, ongoing MS treatment and results from clinical MRI examinations in the form of BPF, number of $T_2$ weighted lesions categorised as 0, 1-9, 10-20 or >20 as well as information on number of GBCA-enhancing and new $T_2$ weighted lesions compared to the most recent, previous MRI.

Inclusion criteria for the study were:
1. Age over 18 years old.
2. Included in the Swedish MS Registry.
3. Receiving care for MS at Umeå University Hospital.
4. Having undergone at least one clinical assessment of BPF before 1st of June 2015, with associated EDSS score (see below) and lesion count.
5. SyMap MRI acquisition performed with 4.5mm slice thickness (to facilitate comparison with the healthy volunteers).

The Swedish MS Registry and the medical records of the radiology clinic, Umeå University Hospital, were used to identify patients suitable for inclusion, resulting in 287 individuals for consideration.

Exclusion criteria were:
1. Clinical MS relapse within 30 days prior to the MRI (n=2).
2. Technical problems / severe MRI artefacts (n=7).

One participant was excluded due to not having MS (included in the registry due to previous misdiagnosis). The final number of individuals included with MS was 278. Of these, a total of 163 individuals with at least two MRIs ≥1 year apart were identified. A total of 68 individuals with at least three MRIs were also identified, the first MRI at >1.0 and <1.5 years from baseline and the second >2.0 years from baseline. Group characteristics can be seen in Table III. Individuals with a first symptomatic episode indicative of MS, but without fulfilment of diagnostic criteria (i.e. clinically isolated syndrome, n=12) were categorised as RRMS.

In an effort to adjust for mismatching in timing between MRI and EDSS examination, the EDSS at the time of MRI was approximated from the EDSS assessments made on the closest dates prior to and following each MRI, linearly weighted by the time between EDSS and MRI. Only MRI examinations with at least one EDSS value within 90 days were included. EDSS worsening was defined as previously described by Cree *et al.* (107). To qualify as worsening, a 1.5 point increase was required from baseline EDSS 0, a 1.0 point worsening was required from scores 1.0 to 5.0 and a 0.5 point score was required from score 5.5.
<table>
<thead>
<tr>
<th>Baseline</th>
<th>Three time points</th>
<th>Two time points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Longitudinal</td>
<td>Longitudinal</td>
</tr>
<tr>
<td></td>
<td>N=68 (43 female)</td>
<td>N=164 (119 female)</td>
</tr>
<tr>
<td>Median age (range)</td>
<td>39.5 (21.4-62.5)</td>
<td>40.4 (18.2-70.8)</td>
</tr>
<tr>
<td>Median BPF (IQR)</td>
<td>0.861 (0.050)</td>
<td>0.850 (0.060)</td>
</tr>
<tr>
<td>Mean BPF-residual compared with regression of BPF over age in the healthy controls (SD)</td>
<td>-0.015 (0.044)</td>
<td>-0.019 (0.041)</td>
</tr>
<tr>
<td>Median MS disease duration (IQR)</td>
<td>7.0 (9.6)</td>
<td>6.8 (10.3)</td>
</tr>
<tr>
<td>Median EDSS (IQR)</td>
<td>2.0 (2.0)</td>
<td>2.0 (2.1)</td>
</tr>
<tr>
<td>Mean number of GBCA-L (SD)</td>
<td>0.015 (0.12)</td>
<td>0.018 (0.13)</td>
</tr>
<tr>
<td>Lesion count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-9:</td>
<td>16 (23.5%)</td>
<td>30 (18.3%)</td>
</tr>
<tr>
<td>10-20:</td>
<td>17 (25.0%)</td>
<td>44 (26.8%)</td>
</tr>
<tr>
<td>&gt;20:</td>
<td>35 (51.5%)</td>
<td>90 (54.6%)</td>
</tr>
<tr>
<td>MS type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRMS:</td>
<td>9 (13.2%)</td>
<td>30 (18.4%)</td>
</tr>
<tr>
<td>PPMS:</td>
<td>2 (2.9%)</td>
<td>9 (5.5%)</td>
</tr>
<tr>
<td>Treatment</td>
<td>Baseline</td>
<td>1st Follow-up</td>
</tr>
<tr>
<td>None:</td>
<td>3 (4.4%)</td>
<td>2 (2.9%)</td>
</tr>
<tr>
<td>Injections:</td>
<td>31 (45.6%)</td>
<td>3 (4.4%)</td>
</tr>
<tr>
<td>Infusions:</td>
<td>31 (45.6%)</td>
<td>61 (90%)</td>
</tr>
<tr>
<td>Orals:</td>
<td>0 (0.0%)</td>
<td>1 (1.5%)</td>
</tr>
<tr>
<td>CTX/HSCT:</td>
<td>1 (1.5%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Other:</td>
<td>1 (1.5%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Data missing:</td>
<td>1 (1.5%)</td>
<td>1 (1.5%)</td>
</tr>
</tbody>
</table>

**Follow-up (times counted from baseline)**

| Mean time to 1st follow-up (SD) | 1.2 (0.13) years | 2.6 (0.81) years |
| Mean time to 2nd follow-up (SD) | 2.9 (0.57) years | - |
| Mean annual PBVC at 1st follow-up (SD) | 0.56% (1.0%) | 0.26% (0.64%) |
| Mean ARR at 1st follow-up | 0.093 (0.32) | 0.045 (0.15) |
| Mean number of lesions per year at 1st follow-up | 0.42 (1.50) | 0.23 (1.3) |
| Mean number of new GBCA-L per year at 1st follow-up | 0.27 (1.1) | 0.091 (0.44) |
| Number with worsened EDSS from baseline to 1st follow-up | 10 (15%, 5 RRMS, 4 SPMS, 1 PPMS) | 23 (14%, 10 RRMS, 12 SPMS, 2 PPMS) |
| Number with worsened EDSS from 1st follow-up to 2nd follow-up | 4 (6%, 1 RRMS, 3 SPMS, 2 PPMS) | - |
Table III; Characteristics of the study population in paper IV

The population data for the longitudinal (left page, page number 26) and cross sectional (right page, page number 27) MS cohorts. This table is cited in its entirety and adapted from Table I in Paper IV.

ARR = Annualised relapse rate, BPF = Brain parenchymal fraction, CTX = Cytotoxic drug, EDSS = Expanded disability status scale, GBCA = Gadolinium based contrast agent, GBCA-L = GBCA-enhancing lesions, HSCT = Hematopoietic stem cell transplantation, IQR = Interquartile range, PBVC = Percent brain volume change, PPMS = Primary progressive MS, RRMS = Relapsing remitting MS, SD = Standard deviation, SPMS = Secondary progressive MS

<table>
<thead>
<tr>
<th>One time point</th>
<th>Cross sectional</th>
</tr>
</thead>
<tbody>
<tr>
<td>N=278 (197 female)</td>
<td></td>
</tr>
</tbody>
</table>

**Baseline**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Median age (range)</td>
<td>42.5 (18.2-76.3)</td>
</tr>
<tr>
<td>Median BPF (IQR)</td>
<td>0.846 (0.059)</td>
</tr>
<tr>
<td>Mean BPF-residual compared with regression of BPF over age in the healthy controls (SD)</td>
<td>-0.021 (0.042)</td>
</tr>
<tr>
<td>Median MS disease duration (IQR)</td>
<td>8.9 (12.5)</td>
</tr>
<tr>
<td>Median EDSS at baseline (IQR)</td>
<td>2.0 (3.3)</td>
</tr>
<tr>
<td>Mean number of GBCA-L (SD)</td>
<td>0.025 (0.18)</td>
</tr>
</tbody>
</table>

**Lesion count**

- 0-9: 48 (17.3%)
- 10-20: 64 (23.0%)
- >20: 166 (59.7%)

**MS type**

- RRMS: 202 (72.7%)
- SPMS: 50 (18.0%)
- PPMS: 26 (9.4%)

**Treatment**

- None: 66 (23.7%)
- Injections: 51 (18.3%)
- Infusions: 139 (50.0%)
- Orals: 6 (2.2%)
- CTX/HSCT: 3 (3.8%)
- Other: 11 (14.1%)
- Data missing: 2 (0.72%)
Definition of BPF and Percent brain volume change (PBVC)

The definition of $BPF = \frac{BPV}{ICV}$ was used for papers I, II and IV. For Paper III this definition was the one predominantly found in the studies in a review of the literature, but the definition $BPF = \frac{BPV}{BSC}$ was identified in a small number of studies (53, 108, 109). The definitions are further explained in the introduction section on BPF (page 16-17).

For Paper IV, Percent brain volume change (PBVC) was defined as $(\text{baseline } BPF - \text{endpoint } BPF) / \text{(baseline } BPF)$. A positive PBVC thus indicates brain volume decrease.

MRI

*Conventional $T_1$, $T_2$ weighted and FLAIR images (Paper II)*

Conventional $T_1$, $T_2$ weighted and FLAIR images were acquired on a 3T Philips Achieva MRI scanner (Philips Healthcare, Best, Netherlands). Detailed imaging parameters are further specified in the methods section of Paper II.

For Paper IV, the clinical examinations had all been performed on either a 3T or 1.5T Philips Achieva scanner. For the longitudinal cohorts only individuals having performed all scans on the same 3T scanner were included.

*Synthetic Tissue Mapping (SyMap) (Paper I, II and IV)*

SyMap acquisition was performed with the specialised QRAPMASTER sequence (37) (acquisition time approximately 6 minutes) and post processing was performed with the dedicated software SyMRI Brain Studio 7.0 or 7.2 (SyntheticMR AB, Linköping, Sweden). This method quantifies the $T_1$ and $T_2$ relaxation times and the PD for each voxel in the image. Based on each voxel’s location in $R_1$-$R_2$-PD space (please see introduction section for a more thorough explanation) the boundaries of the ICV (defined as the border between the skull and the CSF where PD is 50% of the PD of bone) as well as the volumes of GM, WM, CSF and NON (tissue not belonging to either three of the other tissue classes) were quantified. The BPV was calculated as $BPV = GM + WM + NON$ and the BPF as $BPF = \frac{BPV}{ICV}$. The automatic SyMap post processing took approximately 1.5 minutes per case.

For Paper IV, only SyMap acquisitions having been performed with 4.5 mm were included, to keep the data technically homogenous and enable comparisons with the healthy volunteers.
**Statistical Parametric Mapping (SPM)**

**Paper II**

Brain segmentations with SPM were performed with the SPM12 toolkit (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) as a module of MATLAB R2014b (The MathWorks Inc., Natick, MA).

The SPM Segment tool was used to process $T_1$ weighted images using default options for the segmentation. The method is based on a unified model of image registration, tissue classification and bias correction (110) and outputs separate brain masks for WM, GM, CSF. These were quantified to determine the volume of each tissue type and BPV and ICV were calculated as $BPV = GM + WM$ and $ICV = BPV + CSF$ respectively, allowing for the calculation of $BPF = BPV / ICV$. Before the processing of the $T_1$ weighted images, they were co-registered to the tissue probability map used by SPM, in order to ensure identical voxel size of 1.5 mm isotropic voxels as well as identical alignment and spatial orientation. The automatic SPM post processing took approximately 6 minutes per case, not including the export and preparation of the images.

**Voxel-Based Morphometry (VBM)**

**Paper II**

The VBM package version 8 (http://dbm.neuro.uni-jena.de/vbm.html) is included as part of the SPM12 toolbox. It was used to register $T_1$-images to an anatomical template and assign each voxel a tissue class based on voxel intensity. The $T_1$-images used were first resampled to 1.0 mm isotropic voxels and the segmentation was performed using default VBM8 options. After automatic quantification of GM, WM and CSF the BPV could be calculated as $BPV = GM + WM$ and the ICV as $ICV = BPV + CSF$ and thus the BPF were calculated as $BPF = BPV / ICV$. The post processing time was approximately 14 minutes per case for the automatic VBM post processing, not including the export and preparation of the images.

**3DSlicer**

**Paper II**

The software 3DSlicer 4.4.0 (http://www.slicer.org) (111) was used to manually segment MR-images (performed by Mattias Vågberg). By use of local thresholding and free hand correction, MR images can be segmented manually under full visual control. The BPV was segmented from axial FLAIR images and the CSF from axial $T_2$ weighted images. The ICV was calculated as $ICV = BPV + CSF$ and the BPF as $BPF = BPV / ICV$. The post
processing by manual segmentation using 3DSlicer took approximately 45 minutes per case, not including the export of the images.

**Lumbar Puncture and analysis of CSF**

*Lumbar puncture*  
*(Paper I)*

Lumbar puncture was performed according to clinical routine at the neurology clinic, Umeå University Hospital, Sweden. Samples of CSF were collected for analysis of NFL and GFAP levels via ELISA.

**Analysis of the levels of NFL in CSF samples**  
*(Paper I)*

Commercially available NFL ELISA kits (NF-light Assay, UmanDiagnostics AB, Umeå, Sweden) were used for quantification of NFL in CSF. The mean intra- and inter-assay coefficients of variation are, as reported by the manufacturer, 4% and 6% respectively. The lower limit of quantification is stated to be 61 ng/l. The ELISA was performed for each CSF sample both at the UmanDiagnostics laboratory (responsible for the analyses was Niklas Norgren) and at the laboratory of Dr. Jonathan Gilthorpe, Department of Pharmacology and Clinical Neuroscience, Umeå University (by Dr. Ann Dring) yielding two independent quantifications of NFL level for each CSF sample.

**Analysis of level of GFAP in CSF**  
*(Paper I)*

Analysis of GFAP in CSF was performed at the Clinical Neurochemistry Laboratory at the Sahlgrenska Academy (Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden, responsible for the analyses was Prof. Henrik Zetterberg) using an in-house ELISA with and intra- and inter-assay coefficient of variation of 4% and 8%, respectively. The ELISA has been described in detail earlier (112).

**Systematic Review**

The systematic review presented in Paper III was performed in two stages. First an original search in the databases PubMed and Scopus was performed on the 7th of January 2016 (Search String 1 as described below) and then again on the 5th of June 2016 with an additional search string in order to
increase sensitivity (search strings 1 and 2). The results of the two searches were combined and duplicates were removed. The search strings used to identify studies having assessed BPF were specified as:

1. All fields: (“brain parenchymal fraction” OR BPF) AND (“MRI” OR "magnetic resonance imaging" OR “MRT” OR "magnetic resonance tomography" OR “MR” OR “CT” OR "computed tomography")
2. Article title, abstract, keywords: (“brain volume” AND (“fraction” OR “fractional”)) AND (“MRI” OR "magnetic resonance imaging" OR “MRT” OR "magnetic resonance tomography" OR “MR” OR “CT” OR "computed tomography")

A total of 1434 articles were identified. Two additional articles were identified independently. The title and abstract of all articles were examined by two independent reviewers (Mattias Vågberg and Anders Svenningsson). A total of 273 articles were shortlisted as being likely to contain relevant data. Both reviewers then examined the reference tables of these articles and 116 additional articles were identified. The full article texts (n=389) were examined by one of the reviewers (Mattias Vågberg, with consensus discussion regarding any uncertainties) and articles were included for data analysis in accordance with the criteria:

1. Written in English.
2. Having studied a neurologically healthy population.
3. Presenting age of the participants.
4. Presenting data on BPF or data that makes it possible to calculate the BPF.
5. Having used either MRI or CT to measure the BPF (with whole intracranial cavity included).

In cases of suspected or confirmed data duplication, only one study was entered into the review. Figure VIII presents an overview of the literature review process in the form of a flow chart.

Regarding any methodological uncertainties, an attempt was made to contact the corresponding author for clarification (108, 113-118). This was successful in two cases (113, 114) and in the remainder of cases, the methods were interpreted as well as possible from the article text.

Data on BPF, mean age, imaging modality and segmentation method were collected from each included study. If a study presented several subpopulations with complete data for each, the subpopulations were entered individually into the review instead of the data for the whole study population, in order to improve the stratification of BPF by age.
Figure VIII: Overview of the literature review process
The literature review process is illustrated by a flow chart. Two reviewers both examined the results from the initial search and the reference tables from the identified studies. One reviewer examined the results from these steps in full text. Any uncertainties were solved by consensus discussion between the two reviewers. This image is cited in its entirety from Figure 1 in Paper III.
Statistics

The software packages SPSS versions 22 and 23 (2013-2015, IBM Corp. Armonk, NY, US), Microsoft Excel 2013 (2012, Microsoft Corporation, Redmond, WA, US) and R 3.2.5 (2016, R Core Team, Vienna, Austria) were used for all statistical calculations.

Inspection of histograms in conjunction with Shapiro-Wilks test for normality and assessment of skewness was used to determine normality of distribution.

Parametric tests were used for variables with normal distribution.
- Independent t-test was used for tests between one continuous variable and one categorical variable with two categories
- Repeated measures Analysis of variance (ANOVA) (post hoc with Bonferroni correction) with Greenhouse-Geisser correction was used to test for differences between continuous variables and categorical variables with more than two categories

Nonparametric tests were used for statistical tests involving variables with non-normal distribution.
- Spearman’s rho was used for nonparametric bivariate correlation testing.
- Wilcoxon signed rank test was used for nonparametric pairwise comparison of continuous variables and Mann-Whitney U-test was used if the variables were independent.
- Kruskal-Wallis non-parametric ANOVA was used for tests between one continuous variable and one categorical variable with more than two categories.
- Fisher’s exact was used for test between two categorical variables.
- Nonparametric partial rank correlation was performed to investigate correlations between BPF, NFL and GFAP while controlling for age.
- Friedman’s Test (post hoc with Dunn’s test with Bonferroni correction) was used for nonparametric repeated measures ANOVA-testing.

Linear regression with first or second-degree polynomial was used for regression analysis.

Weighted mixed regression was used in paper III to investigate the relationship between age and BPF on the study-by-study level. Each observation was weighted by the inverse of the standard deviation, to account for differences in population size and data spread. The individual
post-processing methods were compared by calculation of the estimated marginal means from the regression model.

$P$-values <0.05 were considered statistically significant. Adjustments to control for familywise error in the setting of multiple comparisons were performed as detailed in specific analyses.


\textbf{Results}

\textbf{Paper I}

\textit{Levels of NFL in CSF}

The mean (±SD) levels of NFL for all participants, analysed independently at the two labs, were 332 (±221) ng/l (UmanDiagnostics lab) and 378 (±209) ng/l (lab of Jonathan Gilthorpe). The NFL values from the two labs were highly correlated (R=0.977, \(p<0.001\)) but the difference between them was statistically significant (\(p<0.001\)).

The mean levels of NFL in CSF determined by the two labs correlated with age (\(r=0.870, p<0.001\)) and with the level of GFAP in CSF (\(r=0.655, p<0.001\)). A graphical representation of the relationship between NFL and age can be seen in Figure 1 in Paper I. Mean values of NFL in CSF, stratified by age, can be seen in Table 1 in Paper I.

\textit{Levels of GFAP in CSF}

The levels of GFAP could only be investigated in 52 samples, as one sample was not available for testing (sample lost during transportation to laboratory). The mean (±SD) level of GFAP in CSF was 421 ng/l and levels correlated with age (\(r=0.595, p<0.001\)).

A graphical representation of the relationship between GFAP and age can be seen in Figure 1 in Paper I. Mean values of GFAP in CSF, stratified by age, can be seen in Table 1 in Paper I.

\textit{BPF}

BPF values correlated with age (R=-0.396, \(p=0.005\)) and with levels of NFL (R=-0.308, \(p=0.035\)) but did not correlate with levels of GFAP (\(p=0.151\)). A graphical representation of the association between NFL and BPF can be seen in Figure 2 in Paper I.

\textit{Partial correlations}

When adjusting correlations by age, using non-parametric partial correlations, only the correlation between NFL and GFAP retained a \(p\)-value below 0.05. The partial correlations can be seen detailed in Table 2 in Paper I.
**Paper II**

BPF was determined for each participant using SyMap, SPM12 and VBM8, and for a subgroup with manual segmentation (Table IV). BPF determined by the automated methods SyMap, SPM12 and VBM8 correlated with BPF determined by manual segmentation (R=0.93, \( p<0.001 \); R=0.56, \( p<0.001 \) and R=0.77, \( p<0.001 \) respectively), but there were significant differences between the four methods (\( p\leq0.001 \) for all comparisons). The variation in differences between each of the automated methods to manual segmentation, over different values of BPF, is presented in Figure 2 in Paper II. BPF correlated with age for all four methods (R=0.77, \( p<0.001 \); R=0.52, \( p<0.001 \) and R=0.63, \( p<0.001 \) for the automated methods and R=0.86, \( p<0.001 \) for the manual reference method) (Figure IX).

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Age (years)</th>
<th>SyMap BPF</th>
<th>VBM8 BPF</th>
<th>SPM12 BPF</th>
<th>Manual seg. BPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 to &lt;30</td>
<td>21</td>
<td>22.8 (4.47)</td>
<td>0.887 (0.037)</td>
<td>0.826 (0.010)</td>
<td>0.826 (0.064)</td>
<td>0.907 (0.033) (( n=9; ) age 24.6; IQR 4.0)</td>
</tr>
<tr>
<td>30 to &lt;40</td>
<td>20</td>
<td>33.9 (3.13)</td>
<td>0.870 (0.032)</td>
<td>0.825 (0.019)</td>
<td>0.816 (0.065)</td>
<td>0.876 (0.025) (( n=9; ) age 34.1; IQR 3.7)</td>
</tr>
<tr>
<td>40 to &lt;50</td>
<td>20</td>
<td>44.1 (4.60)</td>
<td>0.866 (0.029)</td>
<td>0.822 (0.017)</td>
<td>0.800 (0.066)</td>
<td>0.875 (0.041) (( n=9; ) age 45.9; IQR 3.4)</td>
</tr>
<tr>
<td>50 to &lt;60</td>
<td>17</td>
<td>52.6 (6.85)</td>
<td>0.837 (0.033)</td>
<td>0.817 (0.025)</td>
<td>0.806 (0.053)</td>
<td>0.844 (0.037) (( n=11; ) age 51.8; 4.8)</td>
</tr>
<tr>
<td>60 to &lt;70</td>
<td>15</td>
<td>67.3 (5.81)</td>
<td>0.811 (0.041)</td>
<td>0.798 (0.022)</td>
<td>0.752 (0.080)</td>
<td>0.819 (0.016) (( n=9; ) age 68.7; 3.6)</td>
</tr>
<tr>
<td>70 to &lt;80</td>
<td>11</td>
<td>75.3 (4.83)</td>
<td>0.778 (0.071)</td>
<td>0.777 (0.046)</td>
<td>0.737 (0.11)</td>
<td>0.766 (0.11) (( n=7; ) age 75.3; IQR 4.8)</td>
</tr>
<tr>
<td>80+</td>
<td>2</td>
<td>82.7 (4.54)</td>
<td>0.697 (0.0060)</td>
<td>0.739 (0.0024)</td>
<td>0.676 (0.019)</td>
<td>- (( n=0 ))</td>
</tr>
<tr>
<td>Whole group</td>
<td>106</td>
<td>45.7 (28.3)</td>
<td>0.857 (0.064)</td>
<td>0.819 (0.028)</td>
<td>0.795 (0.073)</td>
<td>0.865 (0.069) (( n=54; ) age 49.2; IQR 3.3)</td>
</tr>
</tbody>
</table>

**Table IV: BPF as determined by all four segmentation methods**

BPF is presented for each age group and method of BPF determination, as well as for the whole population. All values are median (IQR) except for the group 80+ where range was used instead of IQR due to \( n=2 \). This table is cited from Table 1 in Paper II (copyright Elsevier, used with permission) and adapted for use in the thesis. Five values of IQR had been miswritten in the original publication and have here been corrected. BPF = Brain parenchymal fraction, SyMap = Synthetic tissue mapping, VBM = Voxel based morphometry, SPM = Statistical parametric mapping, IQR = Interquartile range.
Figure IX: BPF in relation to age
BPF plotted against age for each of the four post processing methods. The solid lines represent quadratic regressions with adjusted $R^2$ values of 0.73, 0.67, 0.66 and 0.37 for manual (A), SyMap (B), VBM8 (C) and SPM12 (D) segmentation. This image is cited in its entirety from Figure 1 in Paper II. (copyright Elsevier, used with permission)
Paper III

Results of the literature search

The BPF values of a total of 9269 individuals were identified in a total of 131 independent healthy populations from 95 publications (Table I in Paper III). Population mean ages spanned from 21.0 to 82.7 years. Seven studies were included in Table 1 in Paper III but excluded from parts, or all of the statistical calculations because of either:

1. The brain segmentation did not encompass the complete intracranial space, for example studies using the Freesurfer method (119), which by default does not include the brain stem in the segmentation process. The Freesurfer studies (43, 71, 120-122) (n=5) were not included in the aggregated statistical calculations apart from the method specific regression (Figure 3 in Paper III).

2. Studies presented median values but not means (46, 115, 123-128) (n=8). This is specified in Table I in Paper II. One study (115) did not specify this, and the data was assumed to be presented as mean values.

3. The BPF was defined as BPF=BPV/BSC instead of BPF=BPV/ICV (i.e. the volume within the brain surface contour instead of the intracranial space was used as denominator) (53, 108, 109) (n=3). This is specified in Table I in Appendix I.

All included studies had used MRI for acquiring the data, but several different methods of post processing were represented. Besides Freesurfer, the methodology of post processing was categorised as SIENAX (129), VBM (130), SyMap (46), SPM (110). The SyMap studies were analysed together with the category ‘other’ due to a low number of SyMap data points.

In studies that used several different methods of segmentation, with data presented for each one, only one method was included in this review. Two articles used several methods and scanner settings but presented the mean values of the combined results (131, 132). One study (133) presented BPF values from two separate scans six days apart, we included the mean of those measurements into the review. One study (134) presented data spread as SE but was assumed to be SD for the purpose of this review as the data spread could not represent SE in relation to the reported range. One study (135) presented an age range of 16 to 70 year of age but was included as a principally adult population on the basis of the mean age and data spread. In one study MRI subgroup age was not presented but was approximated from the whole population (136).
**BPF in relation to age**

The population mean ages and mean BPF values correlated with each other (R=0.41, p<0.001). The best fit for the regression was quadratic.

**BPF in relation to method**

When testing for differences in age-adjusted BPF between the various methods for BPF determination there were significant differences (p≤0.05) for all but two comparisons. The difference between SIENAX and SPM was not significant (p=0.074) and the difference between “Other” and SIENAX did not retain statistical significance when adjusted with Bonferroni-Holm correction. The statistically significant differences in age-adjusted BPF between the methods ranged between 0.043 and 0.17.

**Paper IV**

**BPF in MS in relation to lesion count, MS type and disease duration as well as in relation to BPF in healthy individuals**

The median BPF as well as the mean residual from the healthy BPF regression is specified for each of the three MS-cohorts in Table III (page 26-27). There was a significant difference between the residuals from the MS cohort (n=278) and the healthy controls (p<0.001). The BPF correlated with age (R=-0.46, p<0.001) and disease duration (R=0.178, p=0.023). The change in BPF over time for the cohort with three time points is visualised in Figure X. Baseline BPF (p<0.001) values, but not annual PBVC (p=0.44) were different between MS disease types. Post hoc test revealed differences in BPF between RRMS and SPMS/PPMS but not between SPMS and PPMS.

The BPF was lower in patients with lesion counts >20 compared with patients with 10-20 and 0-9 lesions, but there were no differences between the categories 10-20 and 0-9. The BPF correlated with MS disease duration (R=-0.493, p<0.001) and age (R=-0.469, p<0.001). Baseline BPF was not associated with PBVC during follow-up. There was an inverse relationship between early and late annual PBVC in the cohort with three MRIs (n=68). The annual number of GBCA-enhancing and T2-lesions during baseline to 1st follow-up both correlated negatively with concurrent annual PBVC (R=-0.32, p=0.008 and R=-0.28, p=0.02, respectively) and positively correlated with subsequent annual PBVC (R=0.27, p=0.026 and R=0.32, p=0.008, respectively).
Figure X: BPF and EDSS over time

The BPF (top graph) and EDSS (bottom graph) are plotted against time for all individuals with three MRIs (n=68). The solid lines represent linear regressions and the dashed lines represent 95% prediction intervals. This figure is cited and adapted from Figure I in Paper IV.
**Associations with baseline EDSS and prediction of EDSS worsening**

Baseline EDSS was associated with BPF, BPF residual from the healthy controls, MS duration, lesion count, progressive disease course and age (Table V). The correlation between BPF and EDSS was tested for all available time points for all three cohorts and was significant in all tests ($p<0.001$). Change in EDSS over time for the cohort with three time points can be seen in Figure X.

The regression model for each outcome is detailed in Table VI. The BPF residual was not entered into the model to avoid collinearity with BPF ($R=0.86$, $p=0.001$). All other variables with univariate associations to the outcome were entered as predictors.

The only significant predictor for EDSS worsening in the cohort with two time points was progressive disease course. In the cohort with three time points there were no significant predictors in the regression model. As both progressive disease course and early EDSS worsening had significant univariate associations we explored this further by testing the association between early and late EDSS worsening individually for each MS disease type. Early EDSS worsening was a significant predictor in the SPMS subgroup ($p=0.048$) but not for PPMS or RRMS.
<table>
<thead>
<tr>
<th></th>
<th>EDSS worsening from 1st follow-up to 2nd follow-up (Three time points, N=68)</th>
<th>EDSS worsening from baseline to follow-up (Two time points, N=164)</th>
<th>EDSS (Cross-sectional, N=278)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPF at baseline</td>
<td>-</td>
<td>p=0.40</td>
<td>R=-0.52 (p&lt;0.001)*</td>
</tr>
<tr>
<td>BPF-residual compared to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regression of BPF over age in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the healthy</td>
<td>-</td>
<td>p=0.63</td>
<td>R=-0.38 (p&lt;0.001)*</td>
</tr>
<tr>
<td>Age at baseline</td>
<td>-</td>
<td>p=0.068</td>
<td>R=0.40 (p&lt;0.001)*</td>
</tr>
<tr>
<td>MS duration at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesion count at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of GBCA-enhancing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enhancing lesions at baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDSS at baseline</td>
<td>-</td>
<td>p=0.011*</td>
<td></td>
</tr>
<tr>
<td>Baseline treatment group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progressive MS</td>
<td>p=0.012*</td>
<td>p=0.001*</td>
<td>p&lt;0.001*</td>
</tr>
<tr>
<td>Annual PBVC from baseline to 1st follow-up</td>
<td>p=0.31</td>
<td>p=0.17</td>
<td></td>
</tr>
<tr>
<td>Annualised relapse rate from</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseline to 1st follow-up</td>
<td>p=0.52</td>
<td>p=0.72</td>
<td></td>
</tr>
<tr>
<td>Annual number of new lesions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>from baseline to 1st follow-up</td>
<td>p=0.68</td>
<td>p=0.56</td>
<td></td>
</tr>
<tr>
<td>Annual number of new GBCA-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>enhancing lesions from</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baseline to 1st follow-up</td>
<td>p=0.70</td>
<td>p=0.91</td>
<td></td>
</tr>
<tr>
<td>EDSS worsening at 1st</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>follow-up</td>
<td>p=0.009*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table V: Associations with EDSS and EDSS worsening**

Associations between EDSS worsening and baseline and longitudinal variables for the cohort with three time points (left column), two time points (middle column) and one time point (right column). This table is cited in its entirety and adapted from Table II in Paper IV.

* denotes p-value <0.05, BPF = Brain parenchymal fraction, EDSS = Expanded disability status scale, PBVC = Percent brain volume change.
<table>
<thead>
<tr>
<th></th>
<th>EDSS worsening from 1st follow-up to 2nd follow-up (Three time points, N=68)</th>
<th>EDSS worsening from baseline to follow-up (Two time points, N=164)</th>
<th>EDSS (Cross-sectional, N=278)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPF at baseline</td>
<td>-</td>
<td>-</td>
<td>p&lt;0.001*</td>
</tr>
<tr>
<td>Age at baseline</td>
<td>-</td>
<td>-</td>
<td>p=0.11</td>
</tr>
<tr>
<td>MS duration at baseline</td>
<td>-</td>
<td>p=0.69</td>
<td>p=0.83</td>
</tr>
<tr>
<td>Lesion count 10-20</td>
<td>-</td>
<td>-</td>
<td>p=1.0</td>
</tr>
<tr>
<td>(0-9 as reference)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesion count 20+</td>
<td>-</td>
<td>-</td>
<td>p=0.21</td>
</tr>
<tr>
<td>(0-9 as reference)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Progressive MS</td>
<td>p=0.096</td>
<td>p=0.003*</td>
<td>p&lt;0.001*</td>
</tr>
<tr>
<td>EDSS at baseline</td>
<td>-</td>
<td>p=0.54</td>
<td>-</td>
</tr>
<tr>
<td>EDSS worsening at 1st</td>
<td>p=0.069</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>follow-up</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table VI: Regression models to explain baseline EDSS and EDSS worsening

The final models to explain baseline EDSS and EDSS worsening for the cohort with three time points (left column), two time points (middle column) and one time point (right column). All variables with univariate associations in Table IV were entered into the models, with the exclusion of BPF residual, which was excluded due to collinearity. This table is cited in its entirety from Table III in Paper IV.

* denotes p-value <0.05

BPF = Brain parenchymal fraction
EDSS = Expanded disability status scale
PBVC = Percent brain volume change


Discussion

Summary of the main results

The papers in this thesis have presented data on the BPF in healthy individuals of different ages (Papers I, II, III). The BPF in relation to age was explored both on the level of individual healthy volunteers (Papers I and II) and on the level of population means from a systematic literature search (Paper III). The BPF was seen to decrease with age and this relationship was explained best by a quadratic regression.

The differences in BPF determined by different post-processing methods were investigated (Papers II and III) and found to be of a large enough magnitude to be clearly meaningful in the interpretation of the data.

BPF values from the healthy volunteers were investigated in relation to levels of NFL and GFAP in CSF (Paper I) and levels of NFL and GFAP were associated with each other independently of age. An association between level of NFL and BPF was seen, but this association was not statistically significant if adjusted for the person’s age.

Data from the follow-up of individuals with MS were investigated to establish if EDSS worsening could be predicted from clinical variables (Paper IV). BPF correlated with EDSS, consistent with previous studies (55, 137). Of the predictors investigated, only progressive disease course was significantly associated with future EDSS worsening.

The relationships between NFL and GFAP levels and BPF

In the investigation of the relationship between BPF and the two CSF biomarkers NFL and GFAP (Paper I), significant correlations were identified between NFL levels and BPF and between NFL and GFAP levels. All three variables were correlated with age and when correlations were controlled for age, only the correlation between NFL and GFAP retained statistical significance. This could indicate that there is a relationship between the decay of axons and glial cells that is unrelated to an effect of aging. Longitudinal studies with repeated CSF samples from the same individuals would be helpful in elucidating this in more detail. However, this would involve practical difficulties with the recruitment of a large enough cohort of healthy volunteers for repeated lumbar punctures. This is discussed further in the section entitled ‘Limitations of the studies’ (page 48).
The observed correlation between BPF and NFL did not retain statistical significance when age was controlled for. This could potentially be influenced by statistical power, but could also indicate that changes in the two independent parameters are dependent on different underlying biological processes. Another possible explanation is that the observed changes in BPF and NFL levels represent the same underlying biological process, but that the differences in dynamics between the biomarkers makes a cross-sectional setting unsuitable for investigating an association.

Levels of NFL in CSF increase rapidly following an acute CNS injury (138). In contrast, changes in BPF are more likely to represent a cumulative accrual of tissue damage over time. To clearly identify an association between elevated NFL levels and decreased BPF, a longitudinal follow-up with assessments of NFL and BPF at multiple time points would need to be conducted. A subsequent study investigating the relationship between NFL and brain atrophy in MS (139) has identified a significant correlation between the NFL level in CSF at baseline and a change in brain volume over the three subsequent years. This supports an association between elevated NFL levels and brain atrophy. However, since this study only investigated this association in individuals with MS, and not in healthy individuals, it is not possible to generalise these findings to a non-MS disease population.

**BPF in relation to age**

Greater age was associated with lower BPF (papers I-IV). For papers II and III this association was better described with a quadratic regression than with a linear regression, indicating that the rate of change in BPF over age may increase with greater age. This is supported by earlier findings (140) but the reason for this association between rate of BPF change and age is not known. It could be hypothesised that subclinical CNS injury or disease, as may be more prevalent at greater age, could be one factor. However, further research is required to clarify this finding.

It is important to note that a high degree of individual variation can be seen in the BPF values of healthy individuals (Paper II). This indicates that cross-sectional determination of BPF in itself will, in most situations, not support a differentiation between the occurrence or non-occurrence of pathological brain atrophy. A combination of cross-sectional measurements with a longitudinal follow-up would be required.
BPF in relation to post processing method

In Paper II, the BPF of healthy volunteers was determined by three different automated post-processing methods and by manual segmentation for a subgroup. Statistically significant differences were identified between the methods. The largest difference was seen between SyMap and SPM12 (median BPF values of 0.857 and 0.795, respectively), corresponding to a difference of approximately 7.8%. This difference is clearly meaningful if considered in relation to expected change in BPF per year in healthy individuals (approximately 0.2% to 0.7% per year) (48-50). Differences between different post processing methods were further investigated in Paper III by comparing BPF data from studies that had used different MRI segmentation methods. There were statistically significant differences between studies having used different post processing methods, with estimated age-adjusted differences in BPF ranging between 0.043 and 0.17. Analogous with the findings from Paper II, these differences are important if related to expected yearly change in brain volume for healthy individuals.

The findings of differences between different methods for determining BPF are not surprising, as the different post processing methods adopt very different algorithms for image segmentation and volume determination. For example, tissue segmentation using both SPM and VBM includes the comparison with a brain atlas to assist in the classification of the voxels (110). On the other hand, SyMap and other similar methods base the voxel classification solely on the underlying MRI signal. Given these and other differences in available methods for BPF determination, it is not unexpected that differences in the determined BPF values were observed. However, it is very important to consider that such differences exist, especially in the setting of comparing BPF data between different studies. In a comparison of data between studies, as performed in Paper III, it is also important to keep in mind that there are other potential sources of difference between the studies than merely the choice of post processing method, even when adjusting for population age. It cannot be ruled out that hardware parameters, such as scanner type and settings, as well as differences in population characteristics apart from age, could affect the results. However, the differences observed between SyMap, SPM and VBM presented in Paper II indicate that even if scanner, individual and time are controlled for, differences in BPF obtained with different methods are found.

The differences seen between post processing methods highlight the current lack of a gold standard for BPF determination. A consensus gold standard for BPF determination could serve to increase the homogeneity of choice of method between studies as well as simplifying the validation of new
BPF in the MS population and its relationship with EDSS

The mean annual percent brain volume change (PBVC) in the MS cohort that had undergone at least two MRIs was 0.26% per year. To put this into perspective, a recent meta-analysis of treatment effect on brain atrophy in MS has reported an overall annual PBVC of 0.33%/year and 0.27%/year following treatment with first and second line drugs respectively and 0.50%/year following treatment with placebo (62). Due to differences in both methodology and population characteristics, it is difficult to directly compare the PBVC of Paper IV with the PBVC values from the meta-analysis. Despite this, the similarity further indicates that a range of annual PBVC values from approximately 0.25% to 0.35% may be expected in a largely treated MS population. These atrophy rates are similar to what has been reported from healthy individuals (48, 49), but use of different methods for assessment of PBVC impedes direct comparisons.

Baseline BPF was associated with EDSS at all time points, congruent with findings from previous studies (53, 55). In the regression analysis, this association was found to be independent of the other tested variables, which is also previously reported (54, 64).

It is noteworthy that there was an inverse relationship between early and late annual PBVC. The high rate of new and GBCA-enhancing lesions during the time between baseline and 1st follow-up could potentially explain this by the occurrence of inflammatory oedema during the first time period and subsequent pseudoatrophy during the second (88).

Predicting EDSS worsening

For the whole MS group, only progressive disease course was predictive of later EDSS worsening. Figure VIII depicts the change in EDSS over time where considerable short-term variations between measurements can be seen. Considerable variations in EDSS values have been previously described (29, 30) and the clinical significance of short term changes in EDSS have therefore been questioned (30).

In a 10-year prospective study using the same definition of EDSS progression as Paper IV, baseline EDSS was predictive of subsequent EDSS worsening, but early EDSS worsening was not (107). The same study investigated several other clinical variables and also found baseline white matter volume to be
predictive of EDSS worsening. This finding implies an importance of baseline brain atrophy for predicting EDSS worsening, as has also previously been described (141). However, baseline BPF was not found predictive of EDSS worsening in Paper IV, although the follow-up time of Paper IV was considerably shorter.

Results generated in Paper IV indicated that early EDSS worsening was a possible predictor of subsequent EDSS worsening in the SPMS-group but not in the RRMS group. This could potentially be explained by underlying differences in pathophysiology and response to DMTs between the different disease types. Evidence of new disease activity in RRMS may lead to a change in treatment, while progressive worsening in SPMS does not respond to currently available DMTs.

There are conflicting results from previous studies regarding an association between baseline EDSS and risk for subsequent EDSS worsening, although in studies having used different definitions of worsening. A higher EDSS value has been reported to predict a higher (55, 142) risk of reaching the disability milestone of EDSS 6.0, but a lower risk for point based increase in EDSS (107, 143). Baseline EDSS was not found to be predictive of subsequent EDSS worsening in another study (144), similarly to the findings in Paper IV.

Neither baseline BPF nor annual PBVC were found to be predictive of EDSS worsening in Paper IV. Earlier studies have reported a predictive value of both baseline BPF (141) and rate of atrophy (55, 69, 144, 145). Once again, the use of different definitions of EDSS worsening complicates comparisons. This is further exemplified by a study that reported both significant and non-significant predictive values when using two different definitions of EDSS worsening for the same data (69). Previous studies that have reported a prognostic value of brain atrophy assessments have also generally investigated this during longer follow-up periods compared to those used in Paper IV. This could indicate a stronger prognostic value over longer follow-up periods.

**Limitations of the studies**

**Paper I**

The process of lumbar puncture, although associated with very minor medical risks, may be perceived as uncomfortable and frightening, making recruitment of volunteers difficult. For this reason, the sample size of Paper I was small to moderate. This must be kept in mind in the interpretation of the
data both regarding the correlation testing but also regarding the presentation of NFL and GFAP in relation to age. This is especially true for the older age groups, where the number of volunteers were smaller. However, due to the difficulty in recruiting a suitable number of healthy volunteers for lumbar puncture even the data from this moderately sized cohort is of value for understanding the association of these biomarkers with age.

**Paper II**

Two of the segmentation methods used, VBM8 and SPM12, included brain atlas comparisons during the segmentation process. While different atlases can be specified, we chose to use the Montreal Neurological Institute 152 (MNI152) template. This template is formed from the combined MRI findings of a group of 152 young adults (146) but has been reported to perform sub-optimally in an older population (146, 147). The MNI152 template is included in the SPM12 software package and we opted to use it on the basis of it being supplied as the default template. However, it has been suggested that the creation of a template that is specific for the studied population may help to improve the accuracy of the segmentation (146, 147). Regarding the reliance on an external template when using VBM8 and SPM12, it must also be mentioned that in order to facilitate the comparison of the native MRI acquisitions to the template the native acquisitions were first co-registered/resampled to it. It is possible that this co-registration/resampling could have introduced errors into the segmentation process, although the initial examinations to be co-registered were inspected visually both pre- and post-co-registration, without detecting artefacts.

Furthermore, a difference between the VBM8 and SPM12 segmentations to SyMap and the manual segmentation is their reliance on the T1 weighted acquisitions. The similarity in signal intensity between skull and CSF on T1 weighted images could have decreased the accuracy of the VBM8 and SPM12 segmentations.

Various lifestyle factors, for example alcohol consumption (148), have been associated with brain atrophy. Although all the healthy volunteers for papers I, II and IV were interviewed by a research nurse in order to establish the existence of any major health problems, we did not specifically investigate lifestyle factors, which potentially could be one factor in the inter-individual variation that was seen.
It is important to note that the study was a single centre and single scanner study. This means that inter-scanner and inter-centre variations could not be estimated.

**Paper III**

It is important to consider the fact that the ratio of BPV to ICV, while often referred to as BPF, can also be referred to by other names. This could lead to a number of studies not being found by the literature review due to the usage of a different term. Care was taken to minimise this risk by also including studies not specifically using the term BPF and widening the search string to also include articles with the general term “brain volume” in combination with “fraction” or “fractional” in the article title, abstract or keywords. When examining the reference tables, all articles suspected to present the required data were included, regardless of terminology. However, it is probable that some studies that would have been suitable for inclusion have not been identified due to usage of different nomenclature, despite the efforts made.

Furthermore, it is important to note that the databases PubMed and Scopus may not contain all relevant literature for the search. Two major databases for the fields of clinical science and biomedicine are the North American database MEDLINE and the European database EMBASE (149). The combination of PubMed and Scopus cover all MEDLINE and EMBASE citations as well as some sources not indexed in either of these two databases (150, 151). Expanding the literature search to also include the database Web of Science could potentially have yielded additional results. However, a large overlap between Scopus and Web of Science has been reported (152). Considering this overlap together with far better coverage of Scopus in the fields of ‘Biomedical research’ and ‘Natural sciences and engineering’ (152), it is improbable that the inclusion of Web of Science would have been impactful for the conclusions of the study.

An important restriction to the study inclusion was to only include studies that presented complete volumetric data on the whole intracranial cavity, excluding studies that presented only partial data. This decreased the number of possible studies that could be included but increased the stringency and comparability of the data presented. The inclusion of the work by DeCarli *et al.* (153) was an exception as the volumetric data was only based on segmentation of the supratentorial space. We chose to include it in Table 1 (Paper III) due to its large size (2081 individuals) but we excluded it from the aggregated statistical analyses.
A very important limitation of the study was the low number of individuals that experienced EDSS worsening during the follow-up time. This could potentially have been alleviated with a longer follow-up time and a larger population size. However, since this study was based on a geographical population the possibility to affect these factors was limited.

The number of individuals that experienced EDSS worsening was especially low (n=4) in the period between 1st and 2nd follow-up for the cohort with three MRIs. One explanatory factor for this could have been the large number of individuals that switched treatment from no treatment or injections to the generally more efficacious infusion therapies during the period from baseline to 1st follow-up (Table III). The annualised relapse rate (ARR) and number of new lesions per year was higher during this time period, indicating that disease activity could have motivated a treatment switch that could potentially have affected prognosis. It is possible that the selection criteria for the cohort with at least three MRIs introduced a bias towards selecting individuals with more frequent MRIs due to a more active disease.

The retrospective setting is another limitation. An example of this was the need to match EDSS and MRI data by approximating the EDSS score of the MRI date. While this should produce a reasonable surrogate for concomitant MRI and EDSS examinations due to the criteria of at least one EDSS being within 90 days of the MRI, it is not equivalent to direct EDSS measurement.

As the Swedish MS Registry is the main source of data for the study, there is a risk of error due to data entry, as the data has to be entered into the registry manually. A further crosscheck with the medical records at the Neurology Clinic could potentially have increased the data accuracy.

The MRI lesion load is registered as a categorical lesion count scale in the Swedish MS Registry, which has been adapted and used for this study. This means that lesion load data in this study is less precise than an absolute lesion count or determination of total lesion volume. The statistical associations regarding lesion load could potentially have been different if the variable had been determined with higher precision.

A recent study has reported that administration of GBCA may affect the measurement of BPF using the SyMap method (154), with approximately 0.01 (±0.005) higher BPF after GBCA. The examinations of individuals with MS in the context of the clinical care program were in a vast majority of cases
performed with administration of GBCA, while the examinations of the healthy controls were all performed without GBCA. The difference in BPF between individuals with MS and healthy volunteers could thus potentially have been more pronounced in the absence of GBCA in both groups.

Another important factor to consider is the methodological and physiological variation affecting the measurement of BPF (155-157). The methodological variation in the determination of BPF using SyMap has been reported to be relatively low (82). However, the data on the individual variation between measurements from Paper IV is exemplified in Figure VIII, where a number of follow up BPF values were higher than the baseline value. Furthermore, changes far beyond what is expected from change due to atrophy could be seen in select cases. An annual PBVC of 2.9% could be seen between baseline and 1st follow-up for one individual in the cohort with at least three MRIs. In contrast, if investigating only individuals with at least 3 years between two MRIs (n=58, data not shown) the maximum annual PBVC seen in the group was 1.5%. These findings are likely caused by a combination of methodological and physiological variation. These variations seem to be less impactful with longer follow-up times. This is congruent with the idea that as the total cumulative change in BPF becomes larger with longer follow-up, the proportional effect of the variation diminishes. While prognostic value of brain atrophy assessment has been reported even in short-term follow-up (158), this indicates that clinical brain atrophy data in MS becomes more meaningful with longer follow-up and that conclusions in the individual case must be based on several years of follow-up.
How should the clinical follow-up of MS be designed?

Paper IV investigates the ability to predict EDSS worsening in MS by use of variables collected through clinical follow-up. The follow-up of MS, including which variables to follow, varies between caregivers (Personal communication from representatives of different Swedish MS care centres). The differences between caregivers may be related both to local availability of personnel and resources and to different interpretations of the available evidence regarding best practice for follow-ups.

To discuss follow-up in MS and which variables that should be followed, the most meaningful end-point variables need to be established.

**Quality of Life (QoL)**

It could be argued that the most important clinical endpoint for the individual patient would be the quality of life (QoL).

Even though QoL may be considered as a meaningful endpoint, it is not ideal to use QoL-assessments directly to follow MS as QoL-assessments could be affected both by factors attributable to the disease and to other unrelated variables and life events.

Ideal variables for MS follow-up would be disease specific variables for which changes can be measured effectively and intervened against, whilst still being associated with disease-specific long-term impact on QoL. A hypothetical model of this is presented in Figure XI, illustrating that an important part of the pathophysiology of MS is permanent neuronal damage following CNS inflammation (159). The neuronal damage may cause neurological disability and various MS-associated disabilities have been linked to an impaired QoL (160-162). I propose that MS-associated disability can be considered as the primary driving force behind impaired QoL in MS.
Figure XI: A hypothetical model of the determinants of QoL in MS

The pathophysiological processes underlying MS cause neuronal damage via CNS inflammation. A potential primary neurodegenerative pathway leading to persistent neuronal damage is also possible, but has not been proven. Ongoing inflammatory damage may be identified via clinical relapses, new MRI lesions or subacute increase in NFL levels in CSF. Persistent damage to the CNS – measurable as atrophy – causes persistent neurological disability, which can affect QoL. Short-term remitting disability is not covered by this figure. BPF is noted as an example of quantification of atrophy, but other means to quantify atrophy exists. CNS = central nervous system, BPF = brain parenchymal fraction, EDSS = Expanded Disability Status Scale, NFL = neurofilament light polypeptide, QoL = Quality of Life
Clinical disability assessment

EDSS is the most commonly used disability assessment in MS, but as noted in the introduction section (page 6) it has several limitations. While several studies report an association between EDSS scores and QoL (160-162), conflicting results exist (162). This perhaps reflects the fact that QoL may also be independently impacted by MS-associated disabilities that are not easily quantifiable by EDSS, such as cognitive impairment (32, 33) or fatigue (163), or factors not attributable to MS.

The predictive value of EDSS score is limited. A higher baseline EDSS has been associated with a higher risk for reaching the EDSS milestone of 6.0 (55, 142) but a lower risk for point based EDSS worsening from baseline (107, 143). This indicates that choice of definition for EDSS worsening and probably the ordinal nature of the scale in itself are important factors for the reported associations with future disability worsening. Short-term EDSS change does not seem to be predictive of subsequent EDSS outcome (55, 107, 143), but the opposite has been reported (142).

Similar issues exist for other disability assessments as well. As an example, the Paced Auditory Serial Addition Test (PASAT) is a commonly used method for assessment of cognitive impairment in MS. The predictive value of PASAT for future worsening seems limited (107).

Assessing inflammation and neuronal damage

Identification of new lesions on MRI and clinical relapses are the most common methods to identify inflammatory disease activity in clinical follow-up. Quantification of NFL in the CSF is a promising approach, but the need to perform a lumbar puncture to obtain CSF is a major obstacle for widespread implementation in clinical follow-up. This might change as evidence for the feasibility of measuring NFL in serum is emerging (164).

Relapses and MRI lesions

A high baseline ARR, but not ARR during follow-up, has been associated with subsequent 10-year EDSS-outcome (107). Another study saw an association with 10-year EDSS outcome for ARR in early follow-up, but not for ARR during the whole study time (165). A third study reported that ARR during the whole follow-up period was associated with 10-year EDSS outcome.
Occurrences of new inflammatory lesions on T₂ weighted MRI images have been associated with subsequent EDSS worsening (69, 166), but conflicting results exist (107, 142). A study of early lesion formation on late cognitive function also failed to show an association (142). Administering GBCA may enable detection of GBCA enhancing lesions which have been reported to predict future EDSS worsening (167).

Both ARR and MRI lesions are accepted as surrogates for disease activity in clinical trials (168, 169), but their predictive values in clinical follow-up have been questioned (107, 170). It can be seen both in Paper IV and in phase III DMT-studies (9-21) that lesion formation is more frequent than relapses. If considered as markers of new disease activity, lesion formation should thus be the more sensitive measure. However, while a high annual rate of relapses (ARR) may be associated with a decline in QoL (161, 171), this does not seem to be the case for lesion load or formation of new lesions (161).

Atrophy of the CNS

Permanent accrual of disability and increased brain atrophy both represent permanent CNS tissue damage. Brain atrophy could be considered as a more general marker of CNS tissue damage, whilst clinical disability measures, such as EDSS, focus on key pathways that are important for specific neurological functions. Brain atrophy correlates with EDSS (53, 55) and has also been associated with other forms of disability that are not easily quantifiable by EDSS, such as cognitive impairment (54, 107, 142, 172) and fatigue (173). The association with cognitive impairment is more pronounced for atrophy assessments than for lesion load (172). Association between brain atrophy and fatigue have been reported independently of EDSS-score (174). In Paper IV, neither baseline BPF nor PBVC were predictive of EDSS worsening. This is contrasted by earlier studies that have found both baseline BPF (141) and early atrophy rate (55, 69, 144, 145) to be predictive. Furthermore, the treatment effect of DMTs on brain atrophy is associated with the treatment effect on risk of EDSS worsening (59).

As seen in Paper IV as well as in previous studies (54, 60, 64), brain atrophy may be associated with disability independently of lesion load and clinical relapses. This indicates that assessments of brain atrophy provide information that is not gained by assessing either lesion formation or relapse rate.

An association between whole-brain atrophy and QoL has been reported (173). However, another study found no such association (161).
**Measurement variation**

An important aspect in individual follow-up is changes in the variables of interest that are not attributable to an underlying change in the disease.

Considering EDSS, inter-rater variation (29) and day-to-day changes in patient walking distance (28) have been associated with variations of up to 1.5 points. A study of data from 31 clinical MS trials concluded that short-term EDSS worsening may be greatly reliant on measurement variation (30).

Measurement variation in the assessment of brain atrophy was discussed in more detail in the section on limitations of Paper IV (page 52). Both methodological and physiological sources of variation have been reported (82-84, 86). A brain volume change of 1.49% has been reported after technical modifications to the MRI scanner (155). Occasional extreme short-term change in BPF corresponded to an annual PBVC of 2.9% in Paper IV, although the impact of the variation was smaller over longer follow-up.

The identification of new MRI lesions is subject to both intra- and interrater variability (175-177) with up to a threefold difference in lesion count between raters in extreme cases (177). A notable scan to rescan variability has been reported in measurements of lesion volume, higher than that of brain volume (80).

Sources of variability in the assessment of relapses have, to the best of my knowledge, not been studied.

**Difficulties in interpreting the evidence from follow-up studies**

Comparisons with historical controls indicate a more favourable prognosis contemporarily (107), implying that prognosis may be influenced positively by the use of DMTs. This makes it difficult to interpret the prognostic utility of variables from any follow-up study that has not controlled for treatment. A 20-year follow-up study of individuals with a first symptom indicative of MS between 1984 and 1987 reported a stronger association between MRI lesion formation and endpoint EDSS early in the study compared to the later parts of the follow-up (166). As DMTs were not available until the mid-1990s, their introduction could be a contributing factor to this effect. This problem is relevant for available long-term studies (55, 107, 142, 143) and will likely persist since it would be difficult both for practical and ethical reasons to control or randomise treatments during long-term follow-up. Furthermore, the use of different disability endpoints makes comparisons between studies difficult (143).
Summarising the evidence for establishing a follow-up regime

In conclusion, available evidence is either lacking or conflicting regarding several aspects of MS follow-up. Thus, it is necessary to tailor clinical follow-ups as well as possible in relation to available evidence. Both EDSS and disabilities not easily quantified by it may affect QoL but the utility of clinical disability measures to establish a prognosis and guide treatment decisions is limited.

Considering markers of disease activity discussed above, ARR has the most robust association with future EDSS score and perhaps also with long-term impairment of QoL. ARR is a variable that can easily be followed clinically and intervened against with DMTs. A major drawback with ARR is the low sensitivity in a treated population, as compared with lesion formation.

This makes it possible to identify at least three different alternatives to consider for MS follow-up from the perspective of predicting future disability, as a marker for impaired QoL due to MS.

1. To follow ARR and abstain from MRI-examinations after a diagnosis has been made, since all MRI markers of disease activity have been of questionable prognostic value in long-term studies.

2. To follow ARR with the addition of repeated MRI-examinations for identification of new lesions. The rationale for this would be that ARR has low sensitivity for disease activity. Although lesions on MRI are not as robustly linked to disability, there are convincing associations between the treatment effect on lesion formation and the treatment effect on EDSS worsening. This makes lesion formation a valid surrogate marker for following treatment effect that is much more sensitive than ARR. Atrophy is not monitored due to the measurement variation making meaningful interpretation of individual values difficult.

3. As alternative number 2 but with the addition of monitoring CNS atrophy. The rationale for this would be that atrophy assessment adds information about disease activity and neuronal damage that is independent of what is gained by lesion formation, ARR or EDSS. Measurement variation is not unique for brain atrophy but is an important concern in all measurements. As MS is a chronic disease, the long follow-up times of several years that are needed to limit the impact of measurement variation is feasible.

While other combinations of variables for follow-up could be considered, these three alternatives are examples of different approaches to MS follow-up that all hold merit considering the available evidence. Alternative one is
very conservative, and to my knowledge not widely applied. Alternative two is arguably the most common approach to MS follow-up (178, 179), while Umeå University Hospital has adopted alternative three.

The results presented in this thesis cannot answer the question of which of the alternatives that should be used, but it contributes to the combined evidence on which to base this decision. The considerable variation between measurements of BPF supports alternative two. The association between BPF and EDSS, which was stronger than between EDSS and lesion load, and the smaller variations in BPF that was seen over longer follow-up, supports alternative three. We did not have access to QoL-data, and associations with QoL could thus not be studied.

It has to be emphasised that the above reasoning only considers these measurements as tools for predicting future disability that might impact QoL. The examinations discussed may be beneficial in other ways. For example, a neurological examination may detect spasticity which is treatable by symptomatic treatment. Identification of a relapse may lead to treatment with methylprednisolone that could accelerate recovery. Findings from examinations may aid in the understanding of vocational difficulties and enable early rehabilitation or vocational-focused intervention.

As a concluding remark, more research of this subject could help elucidate the aspects of MS follow-up which are now insufficiently known and aid in designing the follow-up to best benefit the individuals suffering from the disease.
The need for future research

An important prospect for future research is to establish a gold standard for atrophy assessment in MS. Ideally this should be quick to perform, easy to incorporate into the standard MRI workflow, require minimal user input and exhibit low scan to rescan variation. SyMap, the method primarily used in this thesis, could be a potential candidate. However, apart from research data supporting a potential gold standard method, an international consensus agreement would be required to provide broad acceptance for the chosen method. This is a considerable hurdle to pass before an international gold standard can be agreed upon. Furthermore, if a consensus gold standard method was available, it would be beneficial to further establish cut-off values for pathological atrophy rates during different lengths of follow-up, to increase the clinical usability. Considering the measurement variation in assessing brain atrophy, specificity of cut-off values could potentially be improved if defined in relation to specific follow-up times and thus taking into account that the relative impact of the variation may be lower during longer follow-up.

Another important area for future study is to better understand measurements of brain atrophy in healthy individuals, especially during long time longitudinal follow-up. In conjunction with this, it is also of importance to better identify the methodological and physiological factors that might affect the assessment of brain atrophy both in the setting of healthy individuals and in MS. Currently, available research has identified some physiological factors of potential importance but as noted in the introduction section there are some conflicting results.

It would also be very valuable to ascertain which endpoint-variables are of the greatest importance in clinical follow-up of MS and how follow-up should be designed in order to give the best opportunities to predict and intervene. This is difficult to accomplish as it would require a large population size, prospectively followed over a long period of time. Furthermore, all changes that could influence the analyses (such as treatment changes) would need to be registered. All disease-specific endpoints of presumed importance would need to be followed, such as physical disability, cognitive impairment, depression, fatigue etc. In conjunction with this, assessments of QoL and cost-effectiveness would be needed. To be able to generalise the results, especially regarding QoL, the study should contain a mixed population regarding various cultural and socioeconomic factors. This would be a costly and complex study to undertake, but the recent prospective 10-year follow-up study by Cree et.al. (107) implies feasibility.
Conclusions

In summary, based on the data presented in papers I to IV, the following conclusions can be drawn:

1. Levels of both NFL and GFAP in CSF were found to be associated with age in healthy individuals. Furthermore, the two CSF biomarkers were found to be inter-correlated even when controlling for age (Paper I).

2. No evident associations between NFL and GFAP to BPF were found in a cross-sectional setting, apart from the influence that the normal aging process had on each of the variables (Paper I).

3. The previously identified association between BPF and age was confirmed (papers II, III and IV).

4. The cross-sectional data implied an association between rate of change of BPF and current age, indicating higher rate of change with higher age (papers II and III).

5. Significant differences exist between different methods for determining BPF from MRI data (papers II and III).

6. A need for a consensus gold standard for BPF determination was highlighted (Paper III).

7. The cross-sectional association between BPF and EDSS that has been previously reported was reproducible within the context of the clinical care program at Umeå University Hospital (Paper IV).

8. A considerable short-term variation exists in estimations of both EDSS and BPF (Paper IV).

9. The ability to predict EDSS worsening in short to medium term clinical follow-up was limited (Paper IV).
Acknowledgements

There are many persons that deserve my thanks and to whom I am grateful, but for the sake of not letting the acknowledgements section fill the majority of the thesis I have limited myself to some special mentions.

I would sincerely like to thank...

...my main supervisor Anders Svenningsson for many things. For inspiration and tutelage. For always having time, despite already being overbooked. For the idea that research meetings can just as well be held in a gym as in an office and for many, many other things.

...my co-supervisor Richard Birgander for his support, for trying to help me understand how an MRI scanner works and for his open, welcoming attitude. During my time as a PhD-student, I have come to the realisation that whatever your mood may be before starting a conversation with Richard, he is likely to improve it.

...my co-supervisor Jonathan Gilthorpe for his support, for his diligent efforts to improve my scientific writing skills, and most of all for taking me under his wings when I took my first trembling steps in the world of research. The path that he started me on is one that still amazes me.

...Umeå University, the Department of Pharmacology and Clinical Neuroscience and especially Professors in Neurology Lars Forsgren, Jan Malm and Peter Andersen and prefect Stig Jacobsson for creating a stimulating environment for research.

...my co-authors Khalid Ambarki, Ann Dring, Gabriel Granåsen, Niklas Norgren, Rasmus Svenningsson, Thomas Lindqvist and Henrik Zetterberg. The work presented in this thesis would of course not have been possible without you.

...Maja and Edvin, for filling my days with joy.

...my family in the north; Agneta, Leif, Daniel and Per and my family in the south; Mia, Janne, Emma and Anna.
Erika Figaro, Karin Forsberg, William Hansson, Carina Olofsson, Jonatan Salzer, Maja-Lena Stålnacke, Anneli Sundström and Peter Sundström for important input on the thesis and most of all for the invaluable help and support that allowed me time to focus on actually writing this thesis despite other important matters requiring attention.

David Bäckström, Elias Johansson and Fredrik Toss for friendship, interesting discussions and critical reading of the thesis.

Jörgen Andersson, Mari Bäckström, Sonja Edvinsson and Carina Gillewård for help and assistance in everything related to lumbar punctures and handling of CSF-samples.

Johan Jacobsson and Hans Lindsten for creating a work environment that not only enables, but actively supports, the combination of research and clinical work.

Birgitta Nordin and Sara Widmark for help and support in everything related to life as a PhD student.

Västerbottens Läns Landsting (VLL), the VLL section for ALF, Neuro Sweden and BiogenIdec AB for financial support to enable the research presented in this thesis.

all my colleagues and co-workers at the Neurology Clinic and the Radiology Clinic for providing an atmosphere of friendliness and respect.

all the participants of the studies.
References


146. Fillmore PT, Phillips-Meek MC, Richards JE. Age-specific MRI brain and head templates for healthy adults from 20 through 89 years of age. Front Aging Neurosci. 2015;7:44.


155. Barkhof F. Brain atrophy measurements should be used to guide therapy monitoring in MS - NO. Mult Scler. 2016.

156. Chard D. Brain atrophy measurements should be used to guide therapy monitoring in MS - Commentary. Mult Scler. 2016.


