

**Computational modelling of imaging markers to support the
diagnosis and monitoring of multiple sclerosis**

Arman Eshaghi

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

of the

University College London (UCL)

2018

I, Arman Eshaghi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Arman Eshaghi

Acknowledgement

I must first thank Alan Thompson for his charisma, which attracted me to multiple sclerosis research when I was an undergraduate student. More importantly, I am grateful for his support and mentorship during my PhD. Without his indispensable support during the difficult early phase of my academic life, I would have never been able to even start my PhD studies at UCL and most probably would have been doing something else somewhere else in the world! I thank Olga Ciccarelli for her support and supervision before and during my PhD. I thank Danny Alexander for his patient guidance throughout my PhD. I am grateful to Frederik Barkhof for his constructive comments which have undoubtedly increased the quality of my work. Alan, Olga, Danny, and Frederik have supervised my PhD and shaped the next stage of my career, and I am grateful for this.

I thank Jeremy Chataway for his support and intellectual input, especially in the MS-STAT clinical trial reported as one of the chapters of this thesis. Dan Altmann has been a great supervisor in the statistical methods from whose knowledge and patience I have learned a lot. I am grateful to Rogier Kievit for his essential help on structural equation modelling and for reading the drafts of the multivariate methods and my fellowship grants. I thank all the members of the Queen Square Multiple Sclerosis Centre including Declan Chard who has always helped me in writing and dealing with challenges in my research. I thank Ferran Prados for his support in image analysis. I thank Wallace Brownlee, Floriana De Angelis, Niamh Cawley, and others who have acquired part of the data presented in this thesis. I thank other members including Jon Steele, Tina Holmes, Charlotte Burt, Marie Braisher, Jon Stutters, David MacManus, William Brown, Carmen Tur, Thalys Charalambous, Karen Chung, Sara Collorone, and Rosanna Cortese. I am grateful to all the members of the Centre for Medical Image Computing (CMIC) especially Alex Young, Raz Marinescu, Nick Firth, and Neil Oxtoby for their help with coding the computational models. I thank Karl Friston who has helped me a lot to shape the next phase of my academic career. I thank my colleagues in other MAGNIMS centres around Europe who have contributed data to my thesis, without whose help this work would not have been possible. I thank Jeniffer Nicholas and Chris Frost for their intellectual input.

I thank my girlfriend, Floriana, for her continued support without which I would have never finished the work presented here.

در پایان از پدر و مادر عزیزم، زهرا هدایتی و محمد علی اسحاقی، که همیشه مشوق پیشرفت من بودن تشکر می کنم. از محمد علی صحرائیان که من رو به دنیای ام اس وارد کرد هم متشکرم و هم اینکه بار اولی که از ایشان پرسیدم ام اس مخفف چیست به من نخندید!

Table of Contents

List of Figures	8
Supplementary figures	9
Abstract	10
First-author journal publications resulting from this thesis	11
1. Multiple sclerosis: diagnosis and monitoring	12
1.2. Introduction	12
1.3. Diagnosis: prime time for MRI	12
1.4. Classification (MS phenotypes)	15
1.5. MS disease modifiers phenotype	18
1.6. Remaining questions and motivations for my research	18
1.7. Differential diagnosis	19
1.8. Neuromyelitis Optica (NMO)	19
1.10. Monitoring disease worsening in MS: the role of MRI	21
1.11. Pathology	23
Background: the methods	27
2.1. Summary	27
2.2. Introduction	28
2.3. Computational diagnosis: a classification problem	28
2.4. Measuring the change: longitudinal modelling of neuroimaging data	33
2.5. Time to event models	40
2.6. Probabilistic (Bayesian) and generative models	41
2.7. Conclusion	43
2.8. My contributions	43

3. Grey matter MRI differentiates neuromyelitis optica from multiple sclerosis using random-forest: application of machine learning methods in differential diagnosis	44
3.1. Summary	44
3.2. Introduction	45
3.3. Methods	46
3.4. Results	50
3.5. Discussion	56
4. Deep grey matter volume loss drives disability worsening in multiple sclerosis	61
4.1. Summary	61
4.2. Introduction	63
4.3. Methods	65
4.4. Results	72
4.5. Discussion	84
5. Progression of regional grey matter atrophy in multiple sclerosis	90
5.1. Summary	90
5.2. Introduction	92
5.3. Methods	94
5.4. Results	101
5.5. Discussion	110
6. Brain atrophy mediates the simvastatin effect on disability: a mechanistic modelling study	117
6.1. Summary	117
6.2. Introduction	119
6.3. Material and Methods	121
6.4. Results	128

6.5. Discussion	138
7. Conclusions and future directions	142
7.1. Novel contributions to the field	142
7.2. Future directions	143
Appendix	146
Chapter 3	146
Chapter 4	150
Chapter 5	155
Chapter 6	159
References	164

List of Figures

Figure 1-1. Evolution of the neuroimaging tools and treatment milestones in multiple sclerosis.....	15
Figure 1-2. Schematic presentation of neuro-axonal loss in multiple sclerosis and its effects in a real patient.	25
Figure 2-1. Classification steps for a supervised learning task that differentiates between group 1 and 2.	29
Figure 2-2. A schematic and simplified presentation of a decision tree for a hypothetical patient.....	32
Figure 2-3. Structure of a generative model.....	42
Figure 3-1. Importance of variables in the classifiers differentiating MS, NMO and HCs.....	52
Figure 3-2. Descriptive statistics of the thickness, surface and volume.....	54
Figure 4-1. Imaging analysis pipeline.....	68
Figure 4-2. Baseline volumes, and annual percentage loss of brain regions in clinical phenotypes and healthy controls.	76
Figure 4-3. DGM volume predicts future progression of EDSS.	79
Figure 4-4. Risk of EDSS-progression during follow-up for each Z-score volume loss of the brain regions at baseline (post-hoc analysis).	81
Figure 5-1. Estimating the most likely sequence of atrophy progression.....	95
Figure 5-2. Comparisons of regional volumes between groups.....	102
Figure 5-3. Sequences of atrophy progression and patient staging.	104
Figure 5-4. Regional atrophy and its sequence of progression in all grey matter regions plus brainstem in relapse-onset disease and primary progressive multiple sclerosis.	106
Figure 6-1. Trial profile and available data.....	128
Figure 6-2. Outcomes with a significant treatment effect.....	131
Figure 6-3. Candidate models and their fit measures.....	132
Figure 6-4. Parameter estimates of the most likely model.....	134
Figure 6-5. Atrophy rates in areas with significant ongoing change.	136

Supplementary figures

Supplementary Figure 3-1. Flow diagram for the diagnostic accuracy of the random-forest classifier in people with multiple sclerosis and neuromyelitis optica.	148
Supplementary Figure 3-2. Standardised lesion volume in people with multiple sclerosis and neuromyelitis optica.	149
Supplementary Figure 5-1. Greedy ascent search.....	156
Supplementary Figure 5-2. Positional variance diagram for CIS/relapse-onset MS based on all brain regions.	157
Supplementary Figure 5-3. Positional variance diagram for PPMS based on all brain regions.	158
Supplementary Figure 6-1. Image analysis pipeline.....	163

Abstract

Multiple sclerosis is a leading cause of neurological disability in young adults which affects more than 2.5 million people worldwide. An important substrate of disability accrual is the loss of neurons and connections between them (neurodegeneration) which can be captured by serial brain imaging, especially in the cerebral grey matter. In this thesis in four separate subprojects, I aimed to assess the strength of imaging-derived grey matter volume as a biomarker in the diagnosis, predicting the evolution of multiple sclerosis, and developing a staging system to stratify patients. In total, I retrospectively studied 1701 subjects, of whom 1548 had longitudinal brain imaging data. I used advanced computational models to investigate cross-sectional and longitudinal datasets. In the cross-sectional study, I demonstrated that grey matter volumes could distinguish multiple sclerosis from another demyelinating disorder (neuromyelitis optica) with an accuracy of 74%. In longitudinal studies, I showed that over time the deep grey matter nuclei had the fastest rate of volume loss (up to 1.66% annual loss) across the brain regions in multiple sclerosis. The volume of the deep grey matter was the strongest predictor of disability progression. I found that multiple sclerosis affects different brain areas with a specific temporal order (or sequence) that starts with the deep grey matter nuclei, posterior cingulate cortex, precuneus, and cerebellum. Finally, with multivariate mechanistic and causal modelling, I showed that brain volume loss *causes* disability and cognitive worsening which can be delayed with a potential neuroprotective treatment (simvastatin). This work provides conclusive evidence that grey matter volume loss affects some brain regions more severely, can predict future disability progression, can be used as an outcome measure in phase II clinical trials, and causes clinical and cognitive worsening. This thesis also provides a subject staging system based on which patients can be scored during multiple sclerosis.

First-author journal publications resulting from this thesis

1. Eshaghi, A., Wottschel, V., Cortese, R., Calabrese, M., Sahraian, M.A., Thompson, A.J., Alexander, D.C., Ciccarelli, O., 2016. Gray matter MRI differentiates neuromyelitis optica from multiple sclerosis using random forest. *Neurology* 87, 2463–2470. <https://doi.org/10.1212/WNL.0000000000003395>
2. Eshaghi, A., Prados, F., Brownlee, W., Altmann, D.R., Tur, C., Cardoso, M.J., De Angelis, F., van de Pavert, S.H., Cawley, N., De Stefano, N., Stromillo, M.L., Battaglini, M., Ruggieri, S., Gasperini, C., Filippi, M., Rocca, M.A., Rovira, A., Sastre-Garriga, J., Vrenken, H., Leurs, C.E., Killestein, J., Pirpamer, L., Enzinger, C., Ourselin, S., Wheeler-Kingshott, C.A.M.G., Chard, D., Thompson, A.J., Alexander, D.C., Barkhof, F., Ciccarelli, O., MAGNIMS study group, 2018. Deep grey matter volume loss drives disability worsening in multiple sclerosis. *Annals of Neurology*. <https://doi.org/10.1002/ana.25145>
3. Eshaghi, A., Marinescu, R.V., Young, A.L., Firth, N.C., Prados, F., Jorge Cardoso, M., Tur, C., De Angelis, F., Cawley, N., Brownlee, W.J., De Stefano, N., Laura Stromillo, M., Battaglini, M., Ruggieri, S., Gasperini, C., Filippi, M., Rocca, M.A., Rovira, A., Sastre-Garriga, J., Geurts, J.J.G., Vrenken, H., Wottschel, V., Leurs, C.E., Uitdehaag, B., Pirpamer, L., Enzinger, C., Ourselin, S., Gandini Wheeler-Kingshott, C.A., Chard, D., Thompson, A.J., Barkhof, F., Alexander, D.C., Ciccarelli, O., MAGNIMS study group, 2018. Progression of regional grey matter atrophy in multiple sclerosis. *Brain*. <https://doi.org/10.1093/brain/awy088>

1. Multiple sclerosis: diagnosis and monitoring

"It would be nice if a physician from London, one of these days, were to gallop up hotspur, tether his horse to the gait post and dash in waving a reprieve—the discovery of a cure!"

Diary of a patient with multiple sclerosis – 1917(Dr. T. Jock Murray, 2004)

1.2. Introduction

Multiple sclerosis (MS) is a chronic demyelinating and neurodegenerative disease of the central nervous system. MS affects more than 2.5 million people worldwide (Browne *et al.*, 2014). In many countries, MS is the main cause of non-traumatic disability in young adults. People with MS have different clinical presentations with unpredictable disease courses. While some may remain without a significant disability, others may deteriorate rapidly and become wheelchair-bound within a few years (Confavreux and Vukusic, 2014). Therefore, tools that can monitor, and ultimately predict, the course of MS could have a considerable impact on patient lives and society.

Moreover, the decision to choose and later switch MS treatments is becoming more challenging. USA's Food and Drug Administration has approved 14 disease-modifying treatments for people with MS, which could be used during different stages of MS. To decide on a treatment, neurologists rely on the history of patients, clinical examination, magnetic resonance imaging (MRI) and laboratory results. With the advancement of computer science, tools that may one day support the diagnosis and monitor disease progression in individual patients are becoming a realistic possibility (Matthews, 2015).

1.3. Diagnosis: prime time for MRI

Neurologists are increasingly relying on MRI to diagnose MS. The core components of MS diagnosis are objective evidence of dissemination of brain lesions throughout the central nervous system (dissemination in space) or their dissemination in time.

Early diagnostic criteria (in the 1960s and 1980s) for MS by Schumacher and Poser were based on clinical presentation, neurologic examination and cerebrospinal fluid tests (Schumacher *et al.*, 1965; Poser *et al.*, 1983). In 2001, a group led by Ian McDonald, for the

first time, incorporated MRI into the diagnostic criteria (McDonald *et al.*, 2001). The dissemination of the symptoms or lesions in space or time were the basis for MS diagnosis, which was only possible by multiple MRIs taken over time. In 2010, a group led by Chris Polman simplified the use of MRI to detect dissemination in time and space using only one scan (Polman *et al.*, 2011). In 2017 an international panel led by Alan Thompson revised the 2010 McDonald criteria to speed the diagnosis of MS (the specific changes are explained below) (Thompson *et al.*, 2017).

According to the 2017 revisions of McDonald criteria, the dissemination in space or time is defined as follows:

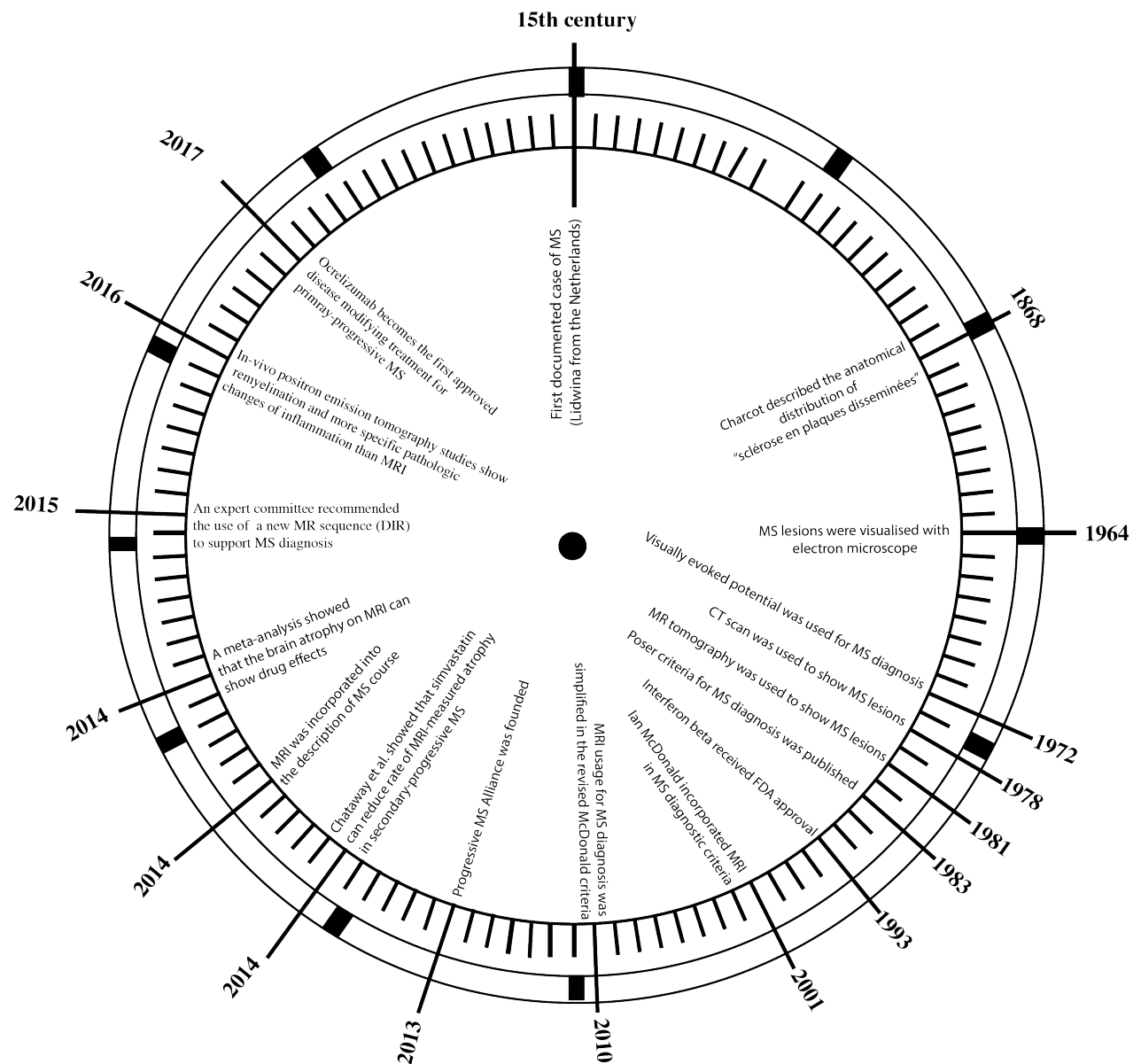
- Dissemination in space: the presence of one or more symptomatic or asymptomatic lesions in at least 2 of 4 specific locations for MS on T2-weighted MRI. Specific lesions for MS are cortical, juxtacortical, periventricular, infratentorial or spinal lesions.
- Dissemination in time: one or both of the following:
 - The appearance of new T2-weighted or gadolinium-enhancing lesions on follow-up MRI. There is no specific interval between baseline and follow-up scans.
 - The simultaneous presence of gadolinium-enhancing and non-enhancing lesions.

In 2014, a meta-analysis found that the effect of drugs on brain atrophy corresponds to those effects on disability (Sormani *et al.*, 2014). During the same year, MRI was incorporated into the definition of MS phenotypes to formally define active and non-active disease (Lublin and Reingold, 1996). The 2010 criteria had only included asymptomatic lesions as the basis for diagnosis. In 2016, the MAGNIMS¹ group gave more value to MRI by including lesions irrespective of symptoms to define dissemination in space and time, in addition to the inclusion

¹ A European committee of MRI experts in MS

of lesions in the optic nerve to define the dissemination of space among other changes (Filippi *et al.*, 2016). Additionally, they suggested the use of advanced MRI (double-inversion recovery), if available, to detect cortical lesions as a sensitive marker for MS (Filippi *et al.*, 2016). In 2017, revisions to 2010 McDonald criteria reinstated the use of cerebrospinal fluid oligoclonal bands in the diagnosis of MS. Moreover, these revisions removed the distinction between symptomatic and asymptomatic lesions for the dissemination in space and time (except for the optic nerve lesions in a patient presenting with optic neuritis). Moreover, other imaging technologies such as optical coherence tomography can now be used as objective clinical evidence in patients who report previous (subjective) visual impairment (Thompson, 2017). Cortical lesions can also be used in the diagnosis (equivalent to juxta-cortical lesions). Therefore, the role of imaging in the diagnosis of MS is expanding rapidly in parallel to the technological advances.

Figure 1-1. Evolution of the neuroimaging tools and treatment milestones in multiple sclerosis.



Abbreviations: DIR; double inversion recovery, MS; multiple sclerosis. MRI; magnetic resonance imaging, FDA; Food and Drug Administration (An American regulatory body for approving medical treatments), CT; computerised tomography.

1.4. Classification (MS phenotypes)

MS is classified into different phenotypes (or subtypes) according to its clinical course (Lublin and Reingold, 1996). Classification is necessary for communication, and more importantly for

clinical decision making based on MS phenotypes. MS is classified into the following phenotypes:

- Clinically isolated syndrome suggestive of MS (will be referred to as clinically isolated syndrome throughout this thesis)
- Relapsing-remitting MS
- Secondary progressive MS
- Primary progressive MS

1.4.1. Clinically isolated syndrome

A clinically isolated syndrome is a single attack of a disease that resembles MS but does not fulfil the MS diagnostic criteria (Miller *et al.*, 2012). Therefore, clinically isolated syndrome has no dissemination in time (monophasic) or no dissemination in space (monofocal) (Lublin and Reingold, 1996). Rarely, patients with clinically isolated syndrome may have clinical evidence for dissemination in time (for example, optic neuritis with Babinski reflex). On the other hand, some people with MS may start their disease without an attack that is suggestive of a demyelinating disease, such as seizures, cognitive problems or encephalopathy (Miller *et al.*, 2012). Majority of patients with the clinically isolated syndrome will convert to MS. The conversion rate of patients with the clinically isolated syndrome to MS differs between studies depending on the geographical location and the presenting symptoms, which ranges from 50% to 80% (Fisniku *et al.*, 2008; Optic Neuritis Study Group, 2008).

1.4.2. Relapsing-remitting MS

People with relapsing-remitting MS constitute 85-90% of all people with MS. People with relapsing-remitting MS have clearly defined attacks. They may completely recover, or their attacks may leave residual deficits. However, there is no gradual progression between the attacks. People with relapsing-remitting MS are either those with the clinically isolated syndrome who have had dissemination in time or those patients who fulfil MRI criteria with the first clinical presentation (Polman *et al.*, 2011). The majority of patients with relapsing-remitting MS enter the progressive stage (secondary progressive MS). The relapsing-remitting stage provides a window of opportunity for treatments, as all the FDA approved disease-modifying

treatments are available for relapsing-remitting MS patients. More recently ocrelizumab has been approved as a disease-modifying treatment for a subgroup of early patients with primary-progressive MS (Hauser *et al.*, 2017). Another oral tablet, siponimod, has shown positive results in the secondary progressive MS (Kappos *et al.*, 2018).

1.4.3. Secondary-progressive MS

If disability in patients with relapsing-remitting MS worsens independent of relapses, patients enter a new phase known as secondary-progressive MS (Lublin and Reingold, 1996). Diagnosis of secondary-progressive MS is retrospective. A study that examined hundreds of definitions for short-term progression against 5-year progression found that *a three-strata progression of Expanded Disability Status Scale confirmed after three months in patients who had a minimum Expanded Disability Status Scale of four with a minimum pyramidal score of two or more* was the best definition for secondary-progressive MS (Lorscheider *et al.*, 2016). Three-strata definition of Expanded Disability Status Scale refers to 1.5-point progression in patients who had a score of 0 before progression, an increase of 1 if the last score is between 1 and 5.5, and an increase of 0.5 if the score was more than 5.5. The majority of patients with relapsing-remitting MS progress to secondary-progressive MS, for example in the Lyon MS cohort after 15 years from relapsing-remitting MS onset, more than half of patients progressed to progressive stage, although the proportion of patients who progress may be less in newer cohorts who have been treated over time with disease-modifying treatment (University of California, San Francisco MS-EPIC Team: *et al.*, 2016). Nonetheless, the effects of disease-modifying treatments in slowing the progression are weak (Lorscheider *et al.*, 2017a). People with secondary-progressive MS may or may not have occasional attacks (relapses).

1.4.4. Primary-progressive MS

People with primary-progressive MS have a gradual progression of disability from the disease onset, with or without relapses. There is no imaging or laboratory examination that can differentiate between different MS phenotypes, and therefore neurologists retrospectively “label” MS phenotypes based on the clinical history. Progressive-relapsing MS was a term used to describe people with a gradual progression from disease onset, who experience

relapses over time(Lublin and Reingold, 1996). However, the new definition of primary-progressive MS includes progressive-relapsing patients, and this term is no longer used. Progressive MS refers to both people with secondary-progressive MS and primary-progressive MS as they show similarities in their disease course (Stys *et al.*, 2012).

1.5. MS disease modifiers phenotype

In 2013's definition of MS disease courses, for the first time, MRI –in addition to clinical history and clinical examination– was incorporated into the description of MS disease phenotypes. For example, patients with the clinically isolated syndrome, relapsing-remitting MS or progressive MS may have “active” or “not active” disease course. Activity is defined as either (1) clinical relapses or, (2) enlarging, or new, T2-weighted MRI lesions, or contrast-enhancing lesions.

International panel also added “progression” to the definition of progressive MS, which is defined by annual clinical assessments. For example, a patient with secondary-progressive MS, with new enhancing lesions but without detectable progression in clinical assessments will have “active secondary-progressive MS without progression” (Lublin and Reingold, 1996).

1.6. Remaining questions and motivations for my research

Although assessment of MRI lesions is part of the diagnosis and the definition for MS phenotypes (e.g., to define active and non-active disease courses), the difference between MS phenotypes using more objective (or quantitative) MRI measures remains unclear. Neuronal cells, especially in the grey matter, degenerate during MS and drive the course of disease toward progression(Jürgens *et al.*, 2016). The loss of grey matter is more strongly associated with clinical and cognitive measures than white matter lesions, or white matter atrophy(Calabrese, Magliozzi, *et al.*, 2015). This ongoing degeneration can be shown, and quantified, by MRI(Bermel and Bakshi, 2006; Geurts *et al.*, 2012). Previous studies on grey matter in people with MS have shown important clues regarding MS diagnosis and progression. For example, some studies have shown more pronounced atrophy in patients with secondary-progressive MS that accelerates over time(Fisher *et al.*, 2008). Others have shown similar rates of atrophy between MS phenotypes(De Stefano *et al.*, 2010). Most of

these studies have been performed on the whole brain, and studies based on longitudinal regional atrophy measures are scarce (Ceccarelli *et al.*, 2008; Riccitelli *et al.*, 2011; Eshaghi *et al.*, 2014; Preziosa *et al.*, 2017). For example, it remains unclear whether a more detailed analysis, including smaller brain regions, could show clinically relevant changes that could be missed by coarse measures of the whole brain— similar to other neurodegenerative disorders (Henneman *et al.*, 2009). Moreover, the grey matter loss is increasingly used to monitor MS progression at the group level in clinical trials (Chataway *et al.*, 2014; De Stefano *et al.*, 2015; Kappos *et al.*, 2015, p. 4; Wang *et al.*, 2016). Therefore, there is an urgent need to further clarify the trajectory of regional volume loss in different MS phenotypes, and in people who progress faster. I will address these questions on grey matter changes in MS phenotypes in **Chapter 4** and **5** of this thesis.

1.7. Differential diagnosis

MS should be differentiated from other central nervous system disorders including vascular, infectious, genetic, and other demyelinating disorders. Some adult central nervous system or systemic diseases that cause white matter lesions in the brain, and may be mistaken for MS are as follows: neuromyelitis optica (NMO), Sjögren syndrome, Lyme disease, systemic lupus erythematosus, CADASIL syndrome (Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy), Behçet disease, polyarteritis nodosa, and sarcoidosis.

One of the most important differential diagnoses of MS is NMO. NMO has a similar presentation to MS and was long considered a variant of MS until the discovery of its cause, anti-aquaporin-4 antibody (NMO-IgG) (Lennon *et al.*, 2004).

1.8. Neuromyelitis Optica (NMO)

NMO is an autoimmune and demyelinating disease of the brain and spinal cord. Optic neuritis and transverse myelitis attacks are the hallmarks of NMO. NMO usually has a relapsing course, and patients may recover or some residual deficits may remain (Wingerchuk, 2007). However, unlike MS, the progressive course is extremely rare (Wingerchuk, 2007; Wingerchuk *et al.*, 2007). Treatment strategies are substantially different between MS and NMO, and even

MS treatments may exacerbate NMO symptoms (Shimizu *et al.*, 2010). Therefore, the differentiation of MS and NMO is of utmost importance in the clinical practice. Diagnosis of NMO is based on the Wingerchuck's criteria (2006), which was later revised in a more recent 2013 international consensus criteria (see **Table 1-1**).

Table 1-1. Diagnostic criteria for neuromyelitis optica in 2006 and its revision in 2013.

Diagnostic criteria	Diagnostic terminology				
Wingerchuck’s 2006 criteria (Wingerchuk <i>et al.</i> , 2006)	Definite NMO	Optic neuritis		1. Contiguous spinal cord lesions that encompass 3 or more vertebral segments 2. Brain MRI not meeting MS criteria 3. Positive test for aquaporin-4 antibody (NMO IgG)	
		Acute myelitis			
		Two of the three supportive criteria			
International consensus criteria (2013)(Wingerchuk <i>et al.</i> , 2015)	NMOSD with AQP4-IgG	At least one core clinical criteria		Core clinical criteria refer to the symptoms that implicate one of these six regions: 1) optic nerve; 2) spinal cord; 3) area postrema of dorsal medulla; 4) brainstem; 5) diencephalon; 6) cerebrum	
		Positive test for NMO-IgG			Preferably with cell-based assay method
		Exclusion of alternative diagnoses			
	NMOSD without AQP4-IgG or with unknown AQP4-IgG	At least two core clinical characteristics that meet the following criteria:		1) One clinical characteristic should be optic neuritis, acute myelitis with LETM or area postrema syndrome; 2) 2 or more different core clinical characteristics; c) fulfilment of additional MRI requirements*	
		Negative tests for AQP4-IgG			
		Exclusion of alternative diagnoses			

* Additional MRI requirement are: 1) acute optic neuritis. Involvement of optic nerve on MRI or only non-specific white matter lesions. 2) acute myelitis: Intramedullary MRI lesions extending more than 3 or more contiguous vertebral segments. 3) area postrema syndrome: Dorsal medulla lesions 4) acute brainstem syndrome: brain stem lesions

Abbreviations: LETM; longitudinally extensive transverse myelitis, AQP4; anti-aquaporin 4 autoantibody, NMOSD; neuromyelitis optica spectrum disorder.

1.9. Remaining questions and motivations for this thesis

Clinical history and examination are the most important tools to differentiate MS from NMO. Brain imaging fulfilling the MRI criteria (explained above) can support those clinical decisions. Advanced neuroimaging studies have further underlined the fundamental differences between NMO and MS. For example, imaging of the cortical lesions in the cortex using an advanced MRI sequence (double-inversion recovery)(Calabrese *et al.*, 2012), ultra-high field MRI(Sinnecker *et al.*, 2012), atrophy measures(Calabrese *et al.*, 2012; Liu, Wang, *et al.*, 2015) and *ex-vivo* studies have shown that cortex is relatively spared in NMO as compared to MS(Saji *et al.*, 2012).

However, the previous studies have focused on MS and NMO characteristics at the group-level. It, therefore, remains unclear whether a more quantitative and automatic approach at the individual level could differentiate MS from NMO. In **Chapter 3** of this thesis, I will present an automatic machine-learning method to measure the brain regional volumes and differentiate NMO from MS in a two-centre setting based on the clinical scans acquired as part of the routine clinical assessment of patients.

1.10. Monitoring disease worsening in MS: the role of MRI

Regular clinical assessment of MS patients is essential to guide treatment strategies. For example, the clinical effectiveness of a prescribed drug is usually assessed by the number of relapses, clinical examination, patient compliance and appearance of new T2-weighted white matter lesions or gadolinium-enhancing lesions on T1-weighted scans during follow-up visits(Brownlee *et al.*, 2017). T2-weighted lesion load in the brain, and especially those of the spinal cord, have been shown to predict conversion from an isolated demyelination attack of the CNS to definite MS(Fisniku *et al.*, 2008; Brown *et al.*, 2017). A study with 20 years of follow-up showed that 79% of patients with the clinically isolated syndrome who had normal brain MRI did not convert to definite MS (Fisniku *et al.*, 2008). Counts or volumes of T2-weighted hyperintense lesions and gadolinium-enhancing lesions on T1-weighted scans have also been used as secondary outcome measures in phase III clinical trials(van Munster and Uitdehaag, 2017). T2-weighted hyperintense lesions represent MRI signal abnormalities that

relate to either inflammation, demyelination and gliosis (or a mixture of these), and as their name implies, appear as hyperintense spots on the T2-weighted MRI scans. Gadolinium-enhancing lesions that appear after injection of the contrast material (gadolinium), on the other hand, represent the breakage of blood-brain barrier (or active inflammation) and appear on T1-weighted MRI sequence. Appearance of gadolinium-enhancing lesions in a single MRI scan or the growth of T2-weighted hyperintense lesions in serial MRI scans, therefore, represents active disease course in MS.

However, the associations between focal lesion volume and clinical outcomes are modest at best and not straightforward, and weaker during progressive MS (Barkhof, 2002). Even in early MS, focal lesions in the white matter can only show the tip of the pathological iceberg. Advanced MRI modalities—which are only used in research studies—can detect abnormalities that are more strongly associated with disability and cognitive measures than focal lesions. Grey matter lesions, for example, are largely missed by the clinical imaging sequences (Kilsdonk *et al.*, 2016). Grey matter atrophy measures from serial brain imaging are usually interpreted as the neuro-axonal loss and are more strongly associated with long-term disability than the focal lesions (Calabrese, Magliozzi, *et al.*, 2015). This damage differs across the cortical areas in the brain; it differentially affects superficial laminar layers and deep gyri of the cortex (Mainero *et al.*, 2015; Haider *et al.*, 2016a). Other advanced modalities such as metabolite and sodium imaging have also shown clinically relevant changes in the brain that are difficult to capture by other methods (Paling *et al.*, 2013; Cawley *et al.*, 2015). However, among all the advanced and quantitative MRI measures, atrophy measures have been more successful as an objective tool to monitor MS worsening, especially during the drug trial (Chataway *et al.*, 2014). This might be due to a wider availability of the T1-weighted MR sequence to measure atrophy, especially for multi-centre trials. In particular, the percentage change of the brain volume is a commonly used measure of atrophy which is calculated from the T1-weighted scans (Smith *et al.*, 2007; Rocca *et al.*, 2017). I will briefly introduce the methods to calculate percentage brain volume changes in **Chapter 2** and apply them in **Chapter 4**.

1.10.1. Limitations of MRI to monitor MS progression

It is important to note that although longitudinal brain volume measures extracted from MRI are useful to detect changes at the group level (e.g., clinical trials) their application at the individual level (e.g., clinical practice) is not proven because they are too variable at the individual level (Barkhof, 2016, 2016). This, in part, is due to the physiological variability in brain volumes which may follow a diurnal pattern (Nakamura *et al.*, 2015) or hormone-dependent changes (Franke *et al.*, 2015). Artefacts and technical variabilities can further increase the measurement error; these include—but are not limited to—head motion (Reuter *et al.*, 2015), inter-scanner variability (Keshavan *et al.*, 2016), and longitudinal scanner drifts (Takao *et al.*, 2011). In short, measures of the percentage brain volume change in serial brain imaging can be used in phase II clinical trials, however, are not ready for individual monitoring of disease progression. As previously mentioned, in the next chapter, I will briefly review available methods to measure atrophy in patients with MS.

1.11. Pathology

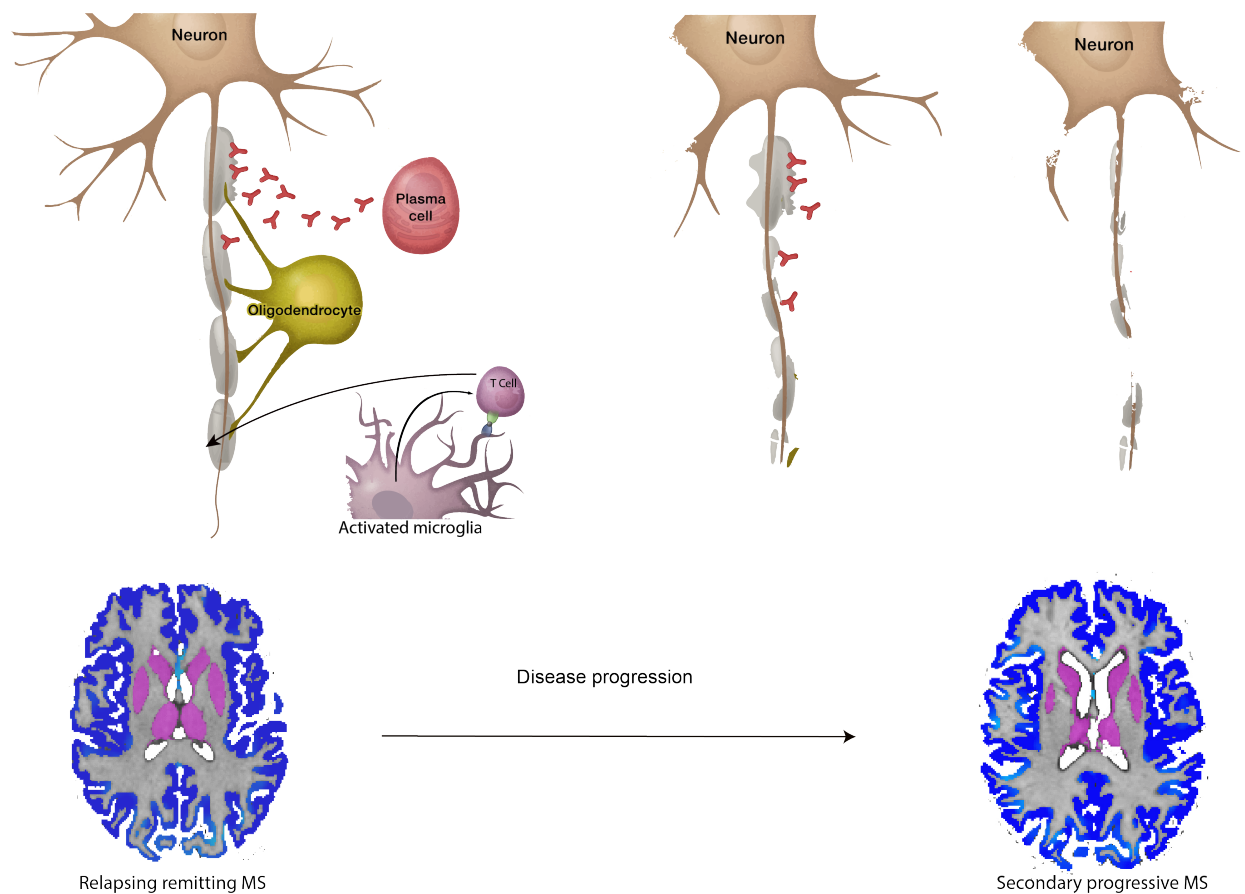
The human brain is one of the most complex phenomenon in the universe. The immune system is also a complex system, which is required for the normal function of any organism. The understanding of MS is, therefore, extremely complex as it ensues from the dysfunction of both. However, regardless of the cause, we know that MS pathology consists of autoimmune attacks on the CNS with demyelination and neurodegeneration. In this brief overview, I will focus on the neuronal loss (or neurodegeneration) and its relation to disability to lay the basis for future chapters (**Figure 1-2**). Loss of neurons, and the connections between them, or neurodegeneration, is the main driver of disability accumulation in MS (Mahad *et al.*, 2015). There are two competing hypotheses about MS pathogenesis. A common one states that aberrant peripheral activation of the immune system starts a pathological cascade with subsequent demyelination and neurodegeneration. An alternative hypothesis assumes that primary death of neuronal cells activates peripheral immune system with subsequent attacks on the CNS. There is evidence in favour of both. However, recent large-scale genome-wide association studies have found several common susceptibility genes among MS and other

autoimmune disorders such as Crohn's disease, ulcerative colitis, psoriasis, rheumatoid arthritis and type-1 diabetes mellitus (International Multiple Sclerosis Genetics Consortium *et al.*, 2011; Baranzini and Oksenberg, 2017). These findings argue against a primary neurodegenerative cause of MS, although they can only explain a proportion (20-30%) of the MS heritability. The exact chain of pathological events in MS may be controversial. However, there is more agreement that innate and adaptive immune dysfunction, chronic activation of microglia, oxidative injury, mitochondrial dysfunction, hypoxia and energy dysregulation all contribute to demyelination, neuro-axonal loss, and disability accumulation at different stages of MS (Mahad *et al.*, 2015). MS affects both the spinal cord and the brain. The grey matter consists of neuronal cell bodies, myelinated and unmyelinated axons and is affected even in early MS (Lucchinetti *et al.*, 2011). Despite the lower total density of myelin in the grey matter than the white matter, the cortical demyelination is extensive (Kutzelnigg *et al.*, 2005). Pathological studies have shown that neurodegeneration is, at least partly, independent of inflammation and demyelination (Carassiti *et al.*, n.d.; DeLuca *et al.*, 2006; Frischer *et al.*, 2009). Based on evidence from imaging and pathological studies I can speculate that a proportion of neuronal death and axonal loss is secondary to inflammation (or focal lesions) in distant areas (Bodini *et al.*, 2016) (such as the white matter or the spinal cord) and also secondary to meningeal inflammation (Magliozzi *et al.*, 2010). Inflammatory milieu imposed by the overlying meninges and cerebrospinal fluid can create a selective neuronal loss in specific neuroanatomical areas, such as those in the deep gyri (the insula) or the deep grey matter areas closer to the ventricles (the thalamus) (Haider *et al.*, 2016b).

The extent of damage in different MS phenotypes is different. For example, a pathological post-mortem study on 29 secondary progressive and seven primary progressive MS subjects showed that the tertiary lymphoid structures could be found in secondary-progressive MS but not primary progressive MS (Magliozzi *et al.*, 2006). This might be due to the longer disease duration in secondary progressive population. However, the pathological differences between MS phenotypes are quantitative rather than qualitative, and therefore, there is no pathological marker to distinguish between MS phenotypes (Mahad *et al.*, 2015).

The main questions that I will try to answer in this thesis will, therefore, be: Do different neuroanatomical areas show different rates of *in vivo* tissue loss? Do different MS phenotypes show differential rates of tissue loss? What is the spatiotemporal pattern of atrophy in patients who have a shorter time to disability progression? Does the increase in the white matter lesion load relate to the grey matter tissue loss?

Figure 1-2. Schematic presentation of neuro-axonal loss in multiple sclerosis and its effects in a real patient.



Neuroaxonal loss or neurodegeneration is thought to be the main substrate of disability progression in multiple sclerosis. The upper row shows that activation of the innate and adaptive immune system attacks myelin, which over the following years ends in neuronal death. This neuroaxonal loss, at the mesoscale, can be seen as brain atrophy. The bottom row shows volume loss in a real patient with relapsing remitting MS (left brain scan), who developed secondary progressive MS (right brain scan) after 13 years. The cortical grey matter is shown in blue, and the deep grey matter is shown in purple.

1.12. Conclusion

In this introductory chapter, I reviewed the definitions of MS phenotypes, the clinical diagnostic criteria and its evolution, and the motivation for using neuroimaging in MS. The evolution of criteria for the diagnosis of MS parallels advances in technology. The latest 2017 revisions of McDonald criteria speed the diagnosis of MS by including symptomatic and non-symptomatic lesions to demonstration dissemination in space and underlines the value of other imaging technologies, such as the optical coherence tomography. Neuronal loss, which is the main driver of disability accumulation, can be captured at the mesoscale as brain atrophy by serial neuroimaging. Therefore, brain atrophy measures may have a high potential in both monitoring disease evolution and prognosticating patient outcomes.

Background: the methods

2.1. Summary

Computational modelling of neuroimaging data is a rapidly evolving field. The methodological advances can enable objective diagnoses of patients and provide more consistent measures of longitudinal change on a larger number of biomarkers than what is possible by a human observer. This chapter focuses on methodological aspects of computational diagnosis and longitudinal models of neuroimaging data. I will briefly review the supervised machine learning methods used in classification problems (e.g., patient diagnosis), the mixed effects longitudinal models to measure the evolution of regional brain volumes, multivariate mechanistic models or structural equation models to formulate scientific hypotheses as statistical models, the event-to-time (survival) models for predicting future progression of disability using baseline imaging biomarkers, and the probabilistic methods for research questions on underlying mechanisms. I will review several alternative methods that are available for specific research questions in this thesis and the rationale for adoption of specific methods. This chapter provides a segue to model applications in the following chapters.

2.2. Introduction

In this chapter, I will briefly introduce the tools to answer the research questions presented in this thesis, all of which relate to diagnosis or monitoring of multiple sclerosis (MS). These questions include the following:

- Can brain volumes distinguish between neuromyelitis optica (NMO) and MS?
- Can brain volumes prognosticate disability outcomes for patients with MS?
- Do rates of brain volume change differ between MS phenotypes?
- Do rates of volume change differ in different neuroanatomical areas?
- How do brain volumes, clinical, and cognitive outcomes relate to each other?

Each following section briefly explains the specific choice of methods in relation to other commonly available tools.

2.3. Computational diagnosis: a classification problem

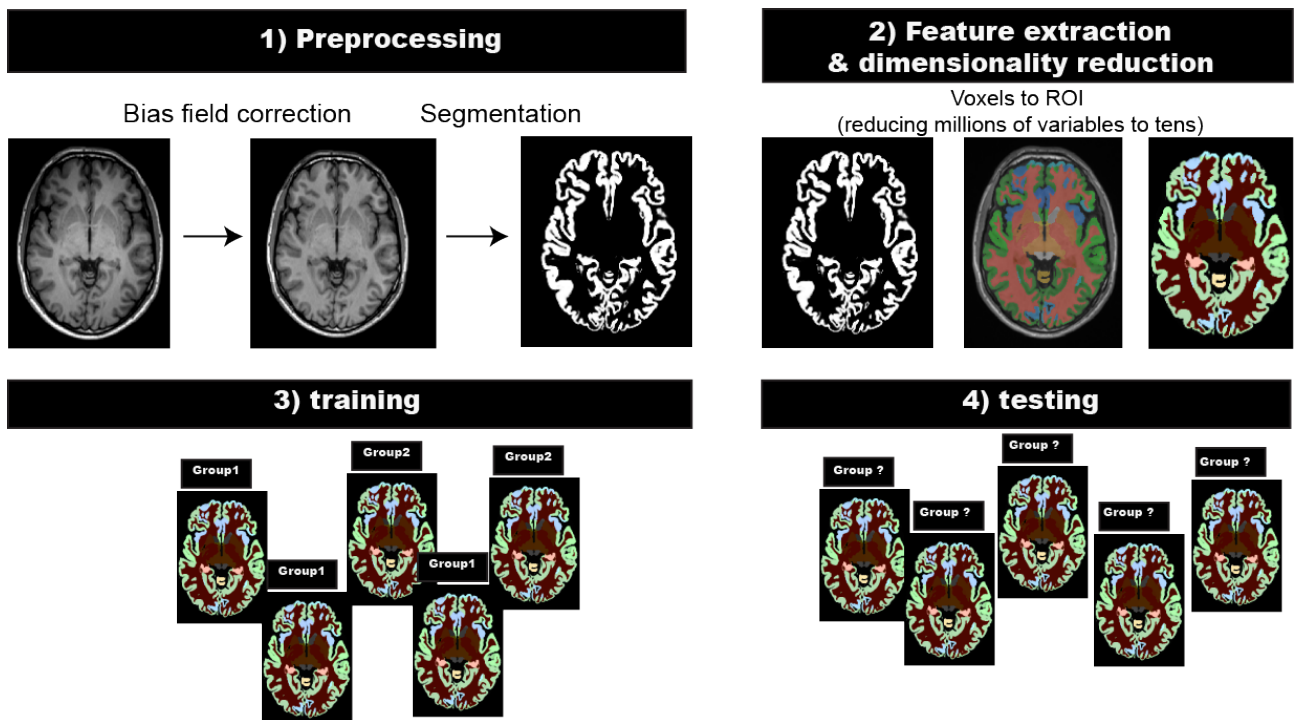
Over the past decade, researchers have started using MRI and machine learning to automatically diagnose (Stefan Klöppel *et al.*, 2008; Klöppel *et al.*, 2012; Stephan and Mathys, 2014; Bendfeldt *et al.*, 2015) or to predict future progression of neurological and psychiatric disorders (Wottschel *et al.*, 2015). Machine-learning techniques can handle a large number of variables in a consistent way, which may not be possible for a human observer. The task of differentiating between two groups is known as classification, which is similar to the case when a neurologist has to differentiate between two similar disorders. In machine-learning classifiers can be used for decision (e.g., predicting diagnosis) or regression problems (e.g., predicting clinical scores) from the MRI data (Wang *et al.*, 2010). The analysis pipeline, irrespective of MRI modality or the algorithm used as the classifier, consists of the following stages:

- 1) **Pre-processing of MRI:** this stage is necessary to reduce the noise in data (e.g., bias field correction, and segmentation), or standardise between subject anatomical variability (e.g., registration to a template). I will explain in detail image processing steps for a classification problem in **Chapter 3**.
- 2) **Feature extraction, and dimensionality reduction:** neuroimaging tools can provide high-dimensional data (e.g., millions of voxels in MRI), which are more than the number

of observations (study participants)(Pereira *et al.*, 2009). Therefore, reducing variables is advantageous for an optimal computation time or to increase the accuracy after removing extraneous information.

- 3) **Training:** the dataset is divided into train and test sets. The train set is used to construct the optimal parameters for the classifier that can best distinguish between groups (decision problem) or provide the closest prediction to the variable of interest (regression problem). There are several methods on how to create test and training sets, or how to deal with an imbalanced number of observations which are discussed elsewhere(Hastie *et al.*, 2011).
- 4) **Testing:** test set, which has been kept separate from the train set until this stage, is used to calculate the unbiased performance of the classifier. **Figure 2-1** shows the classification steps.

Figure 0-1. Classification steps for a supervised learning task that differentiates between group 1 and 2.



Four steps of a classification algorithm are shown in the above figure.

Researchers have used different classifiers in neuroimaging, the most important of which are as follows:

2.3.1. Supervised learning

The outcome variables (e.g., diagnosis for MRI scan), if present, can guide the training process (Hastie *et al.*, 2011). Since the training process uses outcome variables (such as the gold-standard diagnosis from a neurologist) to learn and construct a model, this learning process is called *supervised learning*. Most of the classifiers that are used for differential diagnosis in neurological or psychiatric disorders are supervised classifiers (Stefan Klöppel *et al.*, 2008; Klöppel *et al.*, 2009; Bendfeldt, Klöppel, *et al.*, 2012; Bendfeldt *et al.*, 2015; Wottschel *et al.*, 2015). This classification can distinguish between two groups or more than two groups (multi-class classification). In this thesis, I only focus on the two-class classification – for example, to discriminate between MRI scans of people with MS from those with NMO. Some of the most popular supervised classifiers in neuroimaging include linear models, nearest neighbours, support vector machines, and random forests.

2.3.1.1. *Linear models*

Linear models have been used extensively in neuroimaging. If the decision boundary² In the problem, space is a line (two dimensions), or its surface is a hyperplane (more than two dimensions), the classifier is a linear model. The simplest linear model is a linear regression. The aim here is to construct a line that provides the least error when predicting each observation in the problem space. Mathematically, this could be achieved by minimizing the sums of the squares of the differences between real values and those predicted by the fitted line. This is known as the ordinary least squares. Regression models are easy to interpret and therefore can provide important clues on how the data has been generated. Neuroimaging experiments can provide millions of variables (e.g., voxels). However, only some modifications

² Decision boundary is a line, or a hyperplane in the feature space, that separates two groups from one another.

The classification of observations close to the decision boundary is usually challenging.

of regression models can handle a large number of variables(Tibshirani, 1996; Hastie *et al.*, 2011). Moreover, they are sensitive to the correlation between variables (collinearity) and require homoscedastic data(Hastie *et al.*, 2011). Therefore, they need careful checking of assumptions and selecting variables by an expert.

2.3.1.2. *Nearest neighbours*

Nearest-neighbour methods are non-parametric and model-free methods. Therefore, they are flexible and act as a black-box in classification. To classify an input in the test set, they use those observations in the training set closest to the feature space (or problem space) to the input in the test set. The number of neighbours is usually known as k . Closeness between inputs in the test and training set is a metric, such as Euclidean distance in the feature space. Nearest-neighbours are flexible and can accept a wide variety of inputs. However, because they do not construct a model their results are difficult to interpret. They are used for tasks that interpretation of the model is less important than the final results, such as image segmentation(Steenwijk *et al.*, 2013).

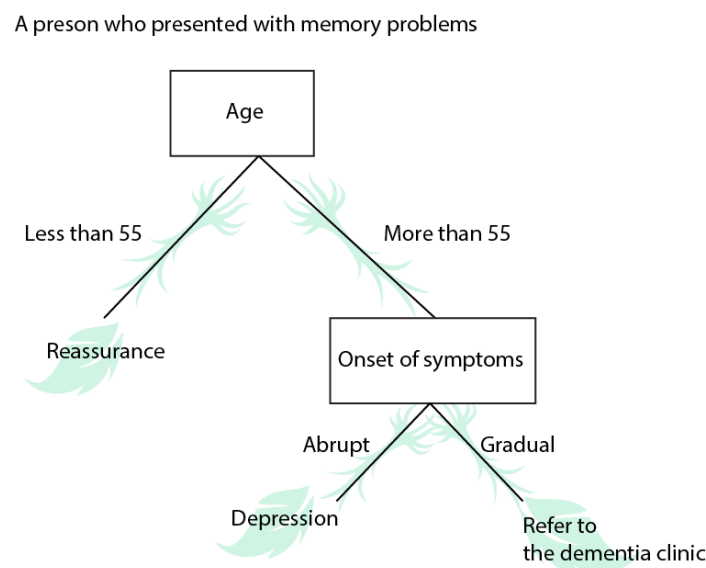
2.3.1.3. *Support vector machine*

Support vector machine (SVM) finds a non-linear boundary in the problem space by searching for a boundary in a transformed (high-dimensional) version of the feature space to separate two groups (Hastie *et al.*, 2011). SVMs have been used extensively in neuroimaging, for example, to differentiate between different Parkinsonian syndromes (Huppertz *et al.*, 2016), predict the future course in people with clinically isolated syndrome (Wottschel *et al.*, 2015), predict the response to treatment in psychiatric disorders (Redlich *et al.*, 2016), and differentiate between benign and non-benign MS (Bendfeldt, Klöppel, *et al.*, 2012). Although SVMs with linear kernels provide easily interpretable results, SVMs with non-linear kernels act as a “black box” and their results may be difficult to interpret.

2.3.1.4. *Random forest*

Flow-charts are used to demonstrate the steps one needs to take to arrive at a decision. Flow-charts are tree-like structures and are known as the decision tree in machine learning (see **Figure 2-2**).

Figure 0-2. A schematic and simplified presentation of a decision tree for a hypothetical patient.



A decision tree (Breiman *et al.*, 1984) consists of the following sections: (1) node; (2) root; (3) leaf. A decision tree takes an input variable, performs a splitting on the value of the input in its node and outputs a target variable (e.g., classification outcome or diagnosis) as its leaf. A decision tree can be constructed by a recursive splitting of the input variables, and comparing the outcome of the leaves with the class that the data belongs to. This recursive splitting continues until it reaches perfect classification or a point where no improvement in its performance is possible. A random-forest classifier is an ensemble method² that uses many decision trees to predict the outcome. Random-forest is a more recent development of decision trees (Breiman, 2001), which reduces the variance of their prediction and improves their generalization. During the training phase, a random-forest classifier constructs many decision trees, but for each decision tree, it uses only a random subset of the input variables (or features). The subset of variables is chosen randomly with bootstrapping (sampling with replacement). This procedure is known as “bagging”, and is necessary to reduce the variance from individual predictions of decision trees, and to reduce the correlation of the decision trees

² Ensemble methods combine results from different classifiers to improve the final prediction.

(Hastie *et al.*, 2011). A random forest can calculate the importance of input features, and therefore the output is easy to interpret. The variable importance is based on the *out-of-bag-error*. The out-of-bag-error for a variable is the mean prediction error of those decision trees that did not have the variable as their features. To calculate the importance of the j th feature, first, the out-of-bag error is calculated for the j th feature. Second, the values of the j th feature are permuted, and again the out-of-bag error is computed. The importance of the j th feature is the difference between these two out-of-bag errors.

Random forests are easier to tune in comparison with other classifiers, are less likely to over-fit⁴, their results are easier to interpret (although they are still a “black box” method)(Hastie *et al.*, 2011), and have shown promise in MRI studies (Eloyan *et al.*, 2012; Lebedev *et al.*, 2014; Moradi *et al.*, 2015). For these reasons, I will use random-forest for automatic differential diagnosis between MS and NMO in the next chapter of this thesis.

2.4. Measuring the change: longitudinal modelling of neuroimaging data

Longitudinal methods in neuroimaging aim to infer trajectory of a biomarker in different populations to inform about an underlying process. Several methods are available for longitudinal analysis of atrophy, which can be divided into 1) image-based methods, 2) statistical methods, and 3) combination methods. Image-based methods are results of image-to-image comparisons, which include boundary-shift integral (BSI), SIENA, and tensor-based morphometry. Statistical methods are more general, such as mixed-effects models, analysis of variance, growth and change models. These are not specific to neuroimaging; they have a longer history and have been borrowed from other fields, such as biostatistics and social sciences. Other authors have used “indirect” and “direct” terms to distinguish between image-based (BSI, SIENA) and statistical methods (Frost *et al.*, 2004). This terminology assumes

⁴ When a statistical model becomes too complex, it models the noise instead of the underlying relationship between variables of interest. This is known as over-fitting. Over-fitted models perform poorly in predicting new cases; they do not generalise well.

that image-based methods subtract values from two visits to provide an *indirect* measure of change, whereas statistical methods rely on the value at each visit, so they are *direct*.

2.4.1. Image-based calculation of atrophy (feature generating methods)

2.4.1.1. *SIENA*

SIENA is a fully automated method to calculate percentage brain volume change between two time-points, which is part of the FSL neuroimaging software (Smith *et al.*, 2002). SIENA performs the following steps to calculate the change: 1) segmentation of the MRI scan into brain and skull, 2) registration and masking in the symmetric space (to avoid bias towards a time point), and 3) change analysis by calculating the motion of brain edge between two time-points. SIENA has been extensively used in phase II and phase III (as the secondary outcome measure) clinical trials in MS. SIENA was designed to calculate the atrophy of the entire brain. However, its recent implementations can calculate percentage volume change in the ventricles (Vrenken *et al.*, 2014).

2.4.1.2. *Boundary shift integral (BSI)*

BSI is very similar to SIENA and consists of the following steps (Freeborough and Fox, 1997): 1) registration of two (or more) time points in a symmetric space, 2) intensity normalisation to adjust for scanner related changes, 3) boundary shift calculation. Some new implementations of BSI perform segmentation as the first step (Leung *et al.*, 2012). BSI has been more popular in Alzheimer's disease field (than MS), in which BSI is the standard to measure brain atrophy changes. BSI and SIENA have an excellent agreement. However, SIENA reports higher percentage atrophy rates (approximately 20% larger atrophy values) (Smith *et al.*, 2007). SIENA and BSI are powerful in measuring global brain volume change, however, are limited in quantifying focal (regional) changes in brain structures and therefore may not be the best option if the aim is to look at regional changes in several brain regions (Hobbs *et al.*, 2009).

2.4.1.3. *Tensor-based morphometry*

This method estimates brain volume changes based on deformation fields that are calculated between two (or more) time points (Ashburner and Friston, 2001; Kipps *et al.*, 2005). The percentage change is calculated by taking the Jacobian determinant of the deformation field.

The promise of tensor-based morphometry is that it does not rely on segmentation, and therefore, can be useful to avoid segmentation related bias and noise. Moreover, it provides voxel-wise changes between two time-points which is advantageous compared to SIENA and BSI. Tensor-based morphometry is, however, limited by its sensitivity to non-biological differences (e.g., choice of brain extraction technique or registration technique) and, therefore, requires careful quality checking.

2.4.1.4. Voxel-based morphometry

Voxel-based morphometry (Ashburner and Friston, 2000) is a segmentation based method that can provide statistical parametric maps (Penny *et al.*, 2011) based on a specific statistical model (e.g., group wise or subject wise). A typical voxel-based morphometry analysis constitutes of the following steps: 1) segmentation of grey and white matter, 2) normalisation of scans to a standard template, 3) smoothing, and 4) applying a statistical model of choice to produce statistical parametric maps. The voxel-based morphometry has been extensively used in neuroimaging and is powerful as an exploratory tool. However, the smoothing step and potential mis-registrations across neuroanatomical regions could reduce its precision. Moreover, the voxel-based morphometry introduces multiple comparisons problem when comparing millions of voxels, like any other method performing many comparisons. Findings from a voxel-based morphometry study, while extremely useful in generating new hypotheses, need to be validated after the discovery phase in a validation cohort.

It is important to note that voxel-based morphometry and other methods that produce statistical parametric maps (e.g., tensor based morphometry or summary statistics methods) can be used to address multiple questions.

The above methods are referred to as **feature-generating methods** because they produce features that can be further analysed by **statistical methods** (see below) to test specific hypotheses.

2.4.2. Statistical calculation of atrophy (methods to analyse generated features)

2.4.2.1. *Summary statistics*

This method has been borrowed from functional MRI studies. Given a set of time-points (or time-series in functional imaging), one can calculate a voxel-wise measure of *change* according to a statistical model for each subject (or MRI session for functional imaging), in what is called a *first-level analysis* (Penny *et al.*, 2011). The results of the first-level analysis, which are statistical parametric maps, are taken to the *second-level* to apply between-group statistical models. Summary-statistics is the standard tool for the analysis of functional images. In the structural imaging, it can be applied to longitudinal data, in a similar vein to the functional imaging. However, a significant limitation is that unlike functional imaging, the longitudinal studies of structural changes cannot provide hundreds or thousands of scans for each subject. Moreover, this method can be biased when there are varying numbers of visits for different subjects, and cannot handle subjects with only one time-point.

2.4.3. Mixed-effects models

The simplest mixed-effects model includes the fixed- and random-effects that can pool individual trajectories together to provide a population trajectory (Fitzmaurice *et al.*, 2011). Fixed-effects represent population changes (population mean trajectory), and random-effects represent individual deviations from the population trajectory (variance of individual trajectories) (Curran *et al.*, 2010). For a linear trajectory, mean intercept (starting point) and the mean slope (rate of change) are the fixed effects; random effects are the estimates of between participant variability. Taken together, for longitudinal MRI analysis, mixed-effects model is an ideal framework to adjust for repeated measures, while estimating group-level trajectories (Bernal-Rusiel *et al.*, 2013). Mixed-effects models are very flexible for “real world” longitudinal data because they are more robust to missing data (that is missed at random) than methods mentioned above. Mixed-effects models can handle varying intervals between visits for different participants, as long as this non-uniformity is random (Bernal-Rusiel *et al.*, 2013). For these reasons, mixed-models are preferred over more classical analysis-of-variance (ANOVA) models for unbalanced repeated-measures studies (most studies are

unbalanced). Despite these strengths, like other regression models, mixed-effects models are designed for “group-level” predictions, and may not work well for individualised predictions (e.g., precision medicine). Moreover, for inference, they can introduce a multiple comparisons problem when applied to millions of voxels, or hundreds of regions.

Mixed-effects models have been extensively used in neuroimaging (Bernal-Rusiel *et al.*, 2013). This method is the standard model in longitudinal cortical thickness analysis in the Freesurfer software package. Moreover, other statistics derived from a mixed-effects model can be useful for other purposes. For example, they have been used to adjust the effects of multiple scanners on longitudinal brain volumes changes in MS (Jones *et al.*, 2013; Biberacher *et al.*, 2016). In **Chapter 4** I will apply mixed effects models to compare rates of volume loss in different brain structures across MS phenotypes.

It is important to note that the above methods can complement each other. Therefore, the above classification of statistical and image-based methods is used only to simplify their understanding. For example, linear mixed-effects models have been used with SIENA to provide rates of atrophy when there are more than two time-points (Frost *et al.*, 2004). Alternatively, voxel-based morphometry can be combined with linear mixed-effects models to provide statistical parametric maps (Penny *et al.*, 2011). Another example is the combination of voxel-based morphometry and tensor-based morphometry to limit the maps of change calculated by tensor-based morphometry to specific structures provided by voxel-based morphometry (Ashburner and Friston, 2001; Eshaghi *et al.*, 2014). **Table 2-1** shows the methods explained here with their advantages and disadvantageous in longitudinal structural imaging.

Table 0-1. Methods to measure longitudinal changes in MRI.

Name	Pros	Cons
Feature-generating methods		
SIENA	Well-established for phase II clinical trials in MS	Limited use for regional atrophy
Boundary shift integral	Well-established for phase II trials in Alzheimer's disease and MS	Limited use for regional atrophy
Tensor-based morphometry	Sensitive method for voxel-wise changes	Sensitive to non-biological changes and requires precise registration
Voxel-based morphometry	Sensitive method optimal for exploratory studies. Can be used with any statistical model (such as mixed-effects models).	Information can be lost in smoothing and requires precise registration to a population template
Methods for analysing features		
Summary statistics	Simple (and powerful) method to provide within-subject statistical parametric maps	Can be biased in unbalanced and smaller number of visits (than a time-series of a functional imaging experiment)
Mixed-effects models	Flexible and powerful for "real-life" data with missing visits, non-uniform intervals between visits, and multiple centres. Can even handle subjects with one visit.	Difficult to predict individual trajectories which differ from group-wise trajectories.

2.4.4. Structural equation models (mechanistic models)

The statistical (and machine learning) models mentioned so far in this chapter are known as univariate methods, in which the outcome variable is only one variable. For example, whether the diagnosis of a patient is multiple sclerosis or another demyelinating disorder (categorical outcome variable) or the speed of walking for a patient (continuous outcome variable). Multivariate models, however, can handle more than one outcome variable. Therefore, a multivariate model can provide a straightforward answer to a more complex hypothesis than what is possible by multiple univariate methods, although multivariate models are more difficult to fit (e.g., require higher sample sizes) and require expertise. The questions I can answer with a multivariate model include: **how** do atrophy rates influence cognitive, disability deterioration, and patient-reported outcomes?

Decomposing information into parts and analysis of the causal chain of interrelations of variables that is realised as a function is called mechanistic modelling. One of the most well-established multivariate models for mechanistic hypotheses is the structural equation modelling which has been extensively applied in social sciences (Kline, 2015). A structural equation model simultaneously fits multiple interrelated regressions between several variables of interest and can be used to compare different hypotheses.

Structural equation modelling consists of the following steps:

- 1) Formulating scientific hypotheses as competing statistical models. Each statistical model can consist of interrelated regression models.
- 2) Comparing the goodness-of-fit of the competing models and selecting the most likely model given data
- 3) To estimate model parameters in the most likely model to provide interpretable associations between model variables

Structural equation models have been extensively used in social sciences and psychology (Kievit *et al.*, 2014). In **Chapter 6** I will apply this method in a clinical trial to deconstruct the observed treatment effects as the statistical pathways linking intermediate variables to

outcomes. This will give insights into underlying mechanisms of drug actions and disease progression.

It is important to distinguish between *multivariable* and *multivariate* methods. All of the univariate methods in this chapter can be multivariable, meaning they can handle more than one predictor (or independent) variable. However, in the literature the term *multivariate* has been interchangeably used for *multivariable*, which is a common mistake (Hidalgo and Goodman, 2013).

2.5. Time to event models

The questions I can answer using time to event models include: do regional atrophy at baseline predict an earlier disability progression? If so, what brain region can better predict earlier disability progression? One of the most popular methods to calculate the risk of event occurrence (e.g., disability progression) and assess the effects of predictors on this risk is Cox regression model or proportional hazards model (Singer and Willett, 2003). Cox regression models are closely related to survival models and have previously been applied to Alzheimer's disease to test the effects of cortical thinning on the risk of conversion from mild cognitive impairment to Alzheimer's disease (Devanand *et al.*, 2012; Sabuncu *et al.*, 2014). Three key terms should be defined in a Cox regression model, which for the research presented in this thesis were the following:

- 1) **Clinical event** in this thesis is defined as the confirmed Expanded Disability Status Scale (EDSS) progression in relation to the last available clinical assessment. A more detailed definition is provided in **Chapter 4**.
- 2) **Metric for time** is year. Atrophy rates and many of the clinical events in MS are expressed as annual occurrences (e.g., annual relapse rate or annual percentage brain volume loss)
- 3) **The beginning of time** is the interval from the study entry. Other potential candidates for setting the beginning of time include years from the symptom onset, years from the disease onset (diagnosis), or time since birth (age). In this thesis, I have chosen the

study entry as the time zero, because the first available brain imaging had been acquired at this time.

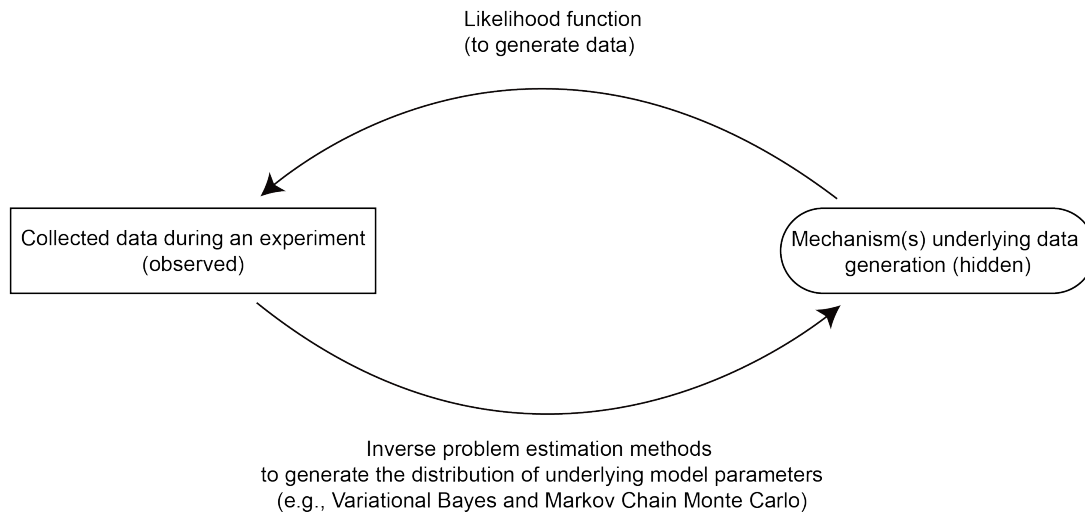
These three elements are necessary for any statistical model that predicts the time to an event.

2.6. Probabilistic (Bayesian) and generative models

The methods introduced so far in this chapter were *deterministic*. A discriminative model provides a value by analysing the observed variables, whereas a probabilistic model can provide a distribution of values. A generative model is (usually) a probabilistic model that generates the distribution of values for a specific phenomenon that are thought to underlie observed variables. Probabilistic and generative models are usually *Bayesian*, meaning that they infer *posterior* probability of a parameter in the model by incorporating the *prior* beliefs with the *observed* variables (Bayes theorem). Bayesian models provide more intuitive answers to research questions; however, they require more expertise, are more computationally expensive, and are not as widely used as deterministic (and frequentist) models. Bayesian models are extremely versatile and are now extensively used in machine learning and artificial intelligence. Some applications of generative models in neuroimaging include image segmentation algorithms of the brain and spinal cord (Blaiotta *et al.*, n.d.), dynamic models of neurophysiological data (Friston *et al.*, 2003, 2017; Daunizeau *et al.*, 2011), and between-domain image translators (Goodfellow *et al.*, 2014; Nie *et al.*, 2016).

Questions that I will try to answer in this thesis by using a generative probabilistic model include: Is there an identifiable order of atrophy progression across different neuroanatomical regions in MS (note the reference to an *underlying* phenomenon)? If so, can I use this order of atrophy progression to stratify patients? Since these questions relate to an *underlying* process that cannot be directly observed, I will use a probabilistic Bayesian model to answer them in **Chapter 5** of this thesis. In the same chapter, I will explain the mathematics underlying a generative model known as *the event-based model* that has previously been used to identify the sequence (or order) of atrophy progression in neuroanatomical areas of patients with Huntington's (Fonteiijn *et al.*, 2012) and Alzheimer's disease (Young *et al.*, 2014).

Figure 0-3. Structure of a generative model.



This figure shows that a generative Bayesian model consists of a likelihood function (forward model), which includes the equations on how to generate the data (the arrow from right to left). For mathematical reasons, there is no equation that can precisely produce the posterior probability of the model. Instead, numerical methods (or approximation methods) can produce the distribution of model parameters in a mathematical problem that is known as an inverse problem. The Markov Chain Monte Carlo method is one of the most powerful methods, however, due to computational expense other methods such as the Variational Bayes might be preferred.

2.7. A note on statistical significance

In this thesis I have used “significance” and “statistical significance” interchangeably. It is important to note that significant *clinical* effects should be interpreted from standardised effect sizes (e.g., Cohen’s *d* reported in Chapter 6) or the magnitude of the difference (e.g., figures in each chapter are designed to compare the magnitude of difference with error bars that represent 95% confidence intervals).

2.8. Conclusion

In this chapter, I briefly reviewed the rationale for choosing specific methods to answer research questions in this thesis. Each method has its advantages and disadvantages, and in many circumstances, a combination of different methods can best answer a specific research question. I will use random forest classification technique in **Chapter 3** because this method requires less parameter tuning. I will use the mixed-effects model in **Chapter 4** because this model is robust to missing and unbalanced longitudinal data. In **Chapter 5** I will use a Bayesian model because the research question relates to the underlying mechanisms that can generate the observed data. In **Chapter 6** I will use a multivariate structural equation model because the research question looks at the potential hypotheses of causal chain of events that result in disability and cognitive score changes (mechanistic hypothesis).

2.9. My contributions

This thesis utilises pre-existing datasets that have been acquired as part of other studies. I have collated and organised these data that constitute imaging data from more than 4000 magnetic resonance scanning sessions from different sources as follows:

- **Chapter 3:** I collated and organised imaging and clinical data from two different imaging centres.
- **Chapter 4 and 5:** I collated and organised imaging data and clinical data from seven MS imaging centres across Europe.
- **Chapter 6:** I collated and organised the imaging, and clinical data for the trial reported in this chapter.

In each of these chapters, I have intellectually conceived ideas behind each chapter, designed all the experiments, performed the experiments, image analyses, statistical analyses, computer coding, design of the figures, writing of the thesis and associated manuscripts. I have used passive verbs where I have used others' work.

3. Grey matter MRI differentiates neuromyelitis optica from multiple sclerosis using random-forest: application of machine learning methods in differential diagnosis

3.1. Summary

I tested whether brain grey matter (GM) imaging measures can differentiate between multiple sclerosis (MS) and neuromyelitis optica (NMO) using random-forest classification. Ninety participants (25 patients with MS, 30 patients with NMO, and 35 healthy controls [HCs]) were studied in Tehran, Iran, and 54 (24 patients with MS, 20 patients with NMO, and 10 HCs) in Padua, Italy. Participants underwent brain T1-weighted and T2-weighted/fluid-attenuated inversion recovery MRI. Volume, thickness, and the surface of 50 cortical GM regions and volumes of the deep GM nuclei were calculated and used to construct three random-forest models to classify patients as either NMO or MS and separate each patient group from HCs. Clinical diagnosis was the gold standard against which the accuracy was calculated. The classifier distinguished patients with MS, who showed greater atrophy especially in deep GM, from those with NMO with an average accuracy of 74% (sensitivity/specificity: 77/72; $p < 0.01$). When I used thalamic volume (the most discriminating GM measure) together with the white matter lesion volume, the accuracy of the classification of MS vs NMO was 80%. The classifications of MS vs HCs and NMO vs HCs achieved higher accuracies (92% and 88%). GM imaging biomarkers, automatically obtained from clinical scans, can be used to distinguish NMO from MS, even in a two-centre setting, and may facilitate the differential diagnosis in clinical practice.

3.2. Introduction

As explained in **Chapter 1** a correct and timely diagnosis of neuromyelitis optica (NMO) is essential for managing patients as treatment options differ considerably from multiple sclerosis (MS). Differential imaging characteristics include: more pronounced brain atrophy, especially in the cortical grey matter (GM) in MS patients than in NMO patients (Calabrese *et al.*, 2012); more severe thalamic changes in MS than NMO (Matthews *et al.*, 2015; Finke *et al.*, 2016); and a lack of cortical lesions in NMO (Calabrese *et al.*, 2012; Sinnecker *et al.*, 2012; Kister *et al.*, 2013).

Machine-learning algorithms have shown promise in classifying MRI scans of patients with neurological disorders (S. Klöppel *et al.*, 2008). Their advantage over human observers is that they can handle a large number of variables from each patient and lack inconsistencies (see **Chapter 2** for further details). Hence, they offer potential in the clinical setting to support the diagnostic process, and have been used successfully in Alzheimer's disease (Lebedev *et al.*, 2014), traumatic brain injury (Lui *et al.*, 2014), and clinically isolated syndromes suggestive of MS (Wottschel *et al.*, 2015).

In this chapter, I tested whether GM measures, obtained from MRI scans acquired as part of clinical protocols, can distinguish MS from NMO. In particular, I calculated the thickness, volume, and surface area of the cortex, and the volume of the basal ganglia, which reflect the underlying GM pathology (Minagar *et al.*, 2013). To test whether this can be used in a multi-centre setting, I used data from two different centres. I paid special attention to the effect of center by explicitly including it in the analysis and then quantifying its contribution to the final results. Finally, I investigated the GM measures that contributed to the discrimination of MS from NMO, to obtain insights into the underlying pathology of these diseases.

3.3. Methods

My primary question was whether imaging biomarkers extracted from routine MRI measures discriminate between MS and NMO. The case-control design of this work provides class-II evidence.

3.3.1. Participants

In this retrospective study, I collected all data from consecutive patients with relapsing-remitting MS and NMO. Participants were recruited in two tertiary centres, one in Iran (January 2009 - December 2012)(Eshaghi *et al.*, 2015) and one in Italy (June 2013 - December 2013): Ninety participants (25 patients with MS, 30 patients with NMO, and 35 healthy controls [HCs]) in Tehran, Iran, and 54 (24 patients with MS, 20 patients with NMO, and 10 HCs) in Padua, Italy. The diagnosis of MS was made according to the McDonald criteria which were revised in 2005 (Polman *et al.*, 2005), and that of NMO according to the Wingerchuk's criteria which were revised in 2006 (Wingerchuk *et al.*, 2006). This study started and was completed before the publication of the new NMO criteria, but a retrospective evaluation showed that all NMO patients fulfilled the 2015 criteria for NMO spectrum disorder (Wingerchuk *et al.*, 2015). Other inclusion criteria were the absences of (1) concomitant neurologic or psychiatric disorders, and (2) clinical relapses or IV-methylprednisolone administration at least six weeks prior to the study entry.

3.3.2. Standard protocol approvals and patient consents

All participants gave written informed consent (Association, 2000), and for each cohort local ethics committee approved the project.

3.3.3. Clinical assessment and MRI protocol

All patients were clinically examined and the Expanded Disability Status Scale (EDSS) was scored on the day of MRI scan (Kurtzke, 1983).

All participants underwent MRI scans that included high-resolution 3D-T1-weighted and T1-weighted/FLAIR imaging, at 1.5T in Padua and 3T in Tehran. The MRI protocol for each cohort is shown in the **Supplementary Table 3-1** in the Appendix.

3.3.4. Image analysis

I analysed the scans of both centres in the same way. The aim of image analysis was to calculate cortical volume, thickness and surface area in 25 cortical regions for each hemisphere defined by LONI Probabilistic Brain Atlas (LPBA) (Shattuck *et al.*, 2008), and the volumes of deep GM nuclei (the accumbens, amygdala, caudate, pallidum, putamen and thalamus). I performed image analysis in ANTs software (v1.9) and FSL (v5.0) (<http://fsl.fmrib.ox.ac.uk/fsl/>). LPBA was chosen as the reference atlas because it is the default atlas used in ANTs.

Hypointense lesions on T1-weighted scans can be erroneously segmented as grey matter. These lesions usually appear as hyperintense lesions on FLAIR/T2-weighted scans. Therefore, I manually constructed binary lesion maps on FLAIR/T2-weighted scans with 3D-Slicer version 4.0, co-registered them with the T1-weighted scan, and used it to fill hypointense lesions of T1-weighted scans to avoid segmentation errors in patients (Sdika and Pelletier, 2009). Next, to reduce the registration errors, I constructed a study-specific template as a common space for registration (Avants *et al.*, 2010). I calculated transformation fields from the atlas space to the common template, and then to each participant's native space. Next I applied the inverse of transformation matrices and deformation fields to LPBA labels to transfer them into native space to extract the imaging measures. To interpolate LPBA labels in the native T1-weighted space, I used nearest neighbor interpolation.

To calculate cortical thickness, I used a new registration based approach that has recently been validated as part of the ANTs software package (Das *et al.*, 2009; Tustison *et al.*, 2014). This method is called Diffeomorphic Registration based Cortical Thickness (DiReCT) estimation, which uses Laplace equation to find one-to-one correspondence between the cortical/white matter and cortical/CSF mantles.

I calculated the surface area, at each cortical region, by dividing cortical volume by the thickness. I used FSL FIRST to calculate volumes of the deep GM structures (Patenaude *et al.*, 2011).

3.3.5. Classification analysis

I performed all statistical analysis in R version 3.1.0 and Scientific Python 2.7 libraries (Scikit-Learn package) (Oliphant, 2007; Abraham *et al.*, 2014). I calculated 157 variables from LPBA atlas regions. Variables were the cortical thickness, surface area, and volume of each cortical region, and the volume of each deep GM nucleus. I divided regional volumes by the total intracranial volume to adjust for different head sizes (Peelle *et al.*, 2012). I adjusted each GM measure separately for the effect of age. I fitted a linear regression line where age was the independent variable and each GM measure was the dependent variable. I calculated the amount of the GM measure that remained unexplained by the regression model (residual of the fit) and used it for all the subsequent analyses.

To investigate the effect of different centres and MRI scanners on the classification, I added the variable “centre” to the analysis. Next, I constructed 3 models to differentiate between each pair of groups: (1) MS vs. NMO; (2) MS vs. HCs; and (3) NMO vs. HCs. For each model, I randomly assigned participants from both centres to either the training or test set, so that each set contained half of the participants. Next, I performed the training step, and then the cross-validation on the left-out half (with 5000 repetitions). The mean and standard deviation of the accuracy of 5000 trained and cross-validated models were calculated. Clinical diagnosis of MS or NMO was the gold standard against which the classification accuracy was calculated. To calculate the statistical significance (or *p*-value) of classification accuracies, I used permutations with 5000 iterations (randomly permuting clinical diagnoses labels), this allows comparison of each classifier with a random classifier. Moreover, to adjust for the effects of centre in calculating statistical significance, I also used a baseline classifier based on the “centre” variable, and compared all the classifiers against this classifier.

I used default values of random forest parameters as provided by the Scikit learn package, and calculated the importance of variables according to the original random-forest algorithm

(Chen *et al.*, 2004). Moreover, when distinguishing between MS and NMO, I built models using cortical region volume, thickness, surface area, and deep GM nuclei volumes, first on their own, and then in combination, to assess the effect of each measure on the accuracy, using a cross-validation approach as explained above. Since a random subset of variables is chosen for each decision tree in a forest of decision trees, random-forest classifiers are not affected by collinearity (correlation between surface, thickness and volume)(Chen *et al.*, 2004). In a *post-hoc* analysis, I simplified the model to assess the accuracy in distinguishing NMO from MS by including the most discriminating GM variable and white matter lesion load. Finally, I trained the NMO vs MS classifier with the scans of NMO patients who were AQP4-positive and tested it to distinguish the scans of NMO patients who were AQP4-negative from those with MS.

To gain a better understanding of GM changes, boxplots of the median and 75th percentile of cortical thickness, surface area and volumes of the subcortical regions were calculated.

3.4. Results

3.4.1. Participants

I included 144 participants; 90 participants were recruited in Tehran, Iran (25 patients with MS, 30 patients with NMO, and 35 HCs) and 54 in Padua, Italy (24 patients with MS, 20 patients with NMO, and 10 HCs). The clinical and demographic characteristics are shown in **Table 3-1** (for treatment information see **Supplementary Table 3-2**).

Table 3-1. Demographic characteristics of participants from each country.

	Tehran cohort			Padua cohort			Both cohorts			
	HC ²	MS ³	NMO ⁴	HC	MS	NMO	HC	MS	NMO	P-value*
Number	35	25	30	10	24	20	45	49	50	
Age (mean ± SD ¹)	32 ± 9.1	32.85 ± 8.5	33.58 ± 10.1	36.04 ± 8.4	36. ± 7.9	42.47 ± 7	33 ± 8.9	34.4 ± 8.3	37 ± 10	0.04
Disease duration (mean ± SD)	–	7 ± 5.5	6.1 ± 3.3	–	8.1 ± 3.2	7.47 ± 2.2	–	7.5 ± 4.4	6.5 ± 3	ns**
Median EDSS (range)	–	2.5 (1-5.5)	3 (1-6)	–	4 (2.5-7)	4.5 (0-7.5)	–	3 (1-7)	3.5(0-7.5)	ns
Gender ratio (female:male)	31:4	22:3	26:4	9:1	12:12	11:9	40:5	34:15	37:13	ns

¹ Standard deviation; ² Healthy controls; ³ Multiple sclerosis; ⁴ Neuromyelitis optica

* P-values are for analysis of variance (ANOVA) for age, T-test for disease duration and EDSS, and Chi-squared test for gender.

** ns = Non-significant

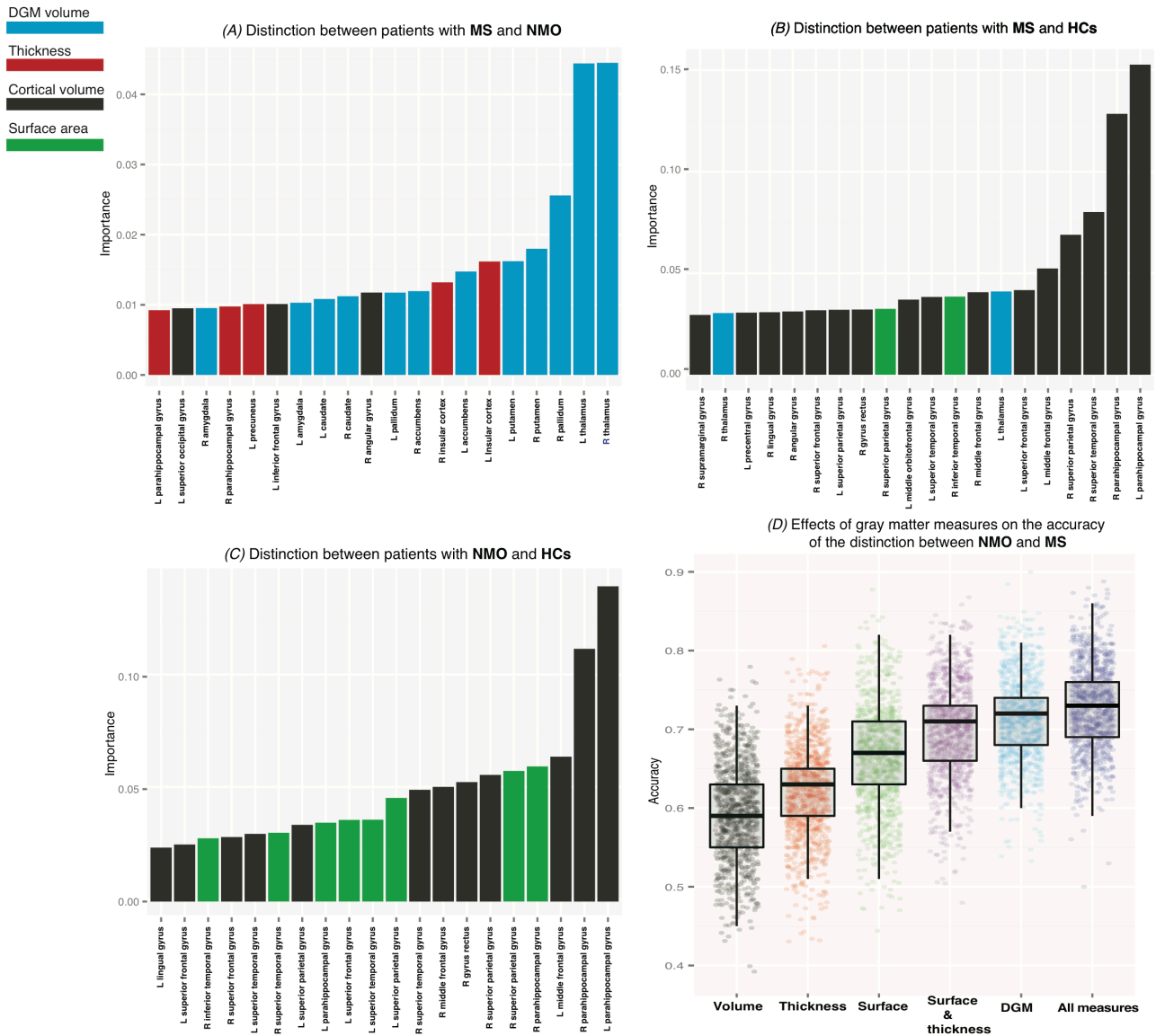
Seventeen out of 30 patients diagnosed with NMO from the Tehran cohort tested positive for aquaporin-4 autoantibody (AQP4-Ab) (immunofluorescence method, <http://www.bioscientia.de/en/>), while 18 out of 20 patients from the Padua cohort were AQP4-Ab positive (cell-based assay, <http://www.euroimmun.com>). Patients in the Italian cohort had a significantly higher disability than those studied in Iran (MS: median EDSS 4.0 vs 2.5, p < 0.05; NMO: 4.5 vs. 3, p<0.05, respectively). Disease duration was not significantly different

between the two cohorts. There was a significant difference in age between groups (see **Table 3-1**).

3.4.2. Classification results

The average accuracy of the model for NMO vs MS trained with all GM imaging features was 74%, with a sensitivity of 77% (i.e., 77% of patients with MS were classified as MS) and a specificity of 72% (i.e., 72% of patients without MS, but with NMO, were classified correctly, see also **Supplementary Figure 3-1**) ($p < 0.01$ calculated against a classifier with randomly changed, or permuted, labels and corrected for multiple comparisons with Bonferroni's method) (see **Table 3-2**). When cortical volume, cortical thickness and surface area were used on their own to classify the two groups of patients, the average accuracy was lower (59%, 62% and 66%, respectively) (see **Figure 3-1**). The results of the analysis remained the same when not adjusting for age.

Figure 3-1. Importance of variables in the classifiers differentiating MS, NMO and HCs.



(A, B, C) show the importance of each variable to classification inside the random-forest classifier. Importance is a relative measure, and is normalized to sum to 1 for each model, and should be used to compare the importance of variables to each model (not among models). (D) shows the accuracy of models with different GM measures, including cortical volumes, cortical surface area, cortical thickness, thickness and surface area in combination, and subcortical volumes. The combination of surface area and thickness obtains a higher accuracy than volume. Each model has been trained and tested 1000 times, after shuffling participants from two centres.

A high accuracy (92%) was obtained when distinguishing MS from HCs (sensitivity/specificity 94/90; $p < 0.001$); the accuracy of classifying NMO vs. HCs was also high (88%; sensitivity/specificity 89/88; $p < 0.001$).

Table 3-2. Classification results.

	Accuracy	Sensitivity	Specificity	Positive	Negative	P
	(average \pm SD¹)	(average \pm SD)	(average \pm SD)	predictive value (average \pm SD)	predictive value (average \pm SD)	value*
Variables: all grey matter measures						
MS ¹ vs NMO ²	74% \pm 5	77% \pm 11	72% \pm 10	73% \pm 8	76% \pm 9	<0.01
MS vs HCs ³	92% \pm 4	94% \pm 6	90% \pm 7	91% \pm 6	93% \pm 7	<0.001
NMO vs HCs	88% \pm 5	89% \pm 8	88% \pm 8	89% \pm 6	87% \pm 8	<0.001
Variables: thalami volume and white matter lesion volume						
MS vs NMO	80% \pm 5	85% \pm 9	76% \pm 9	78% \pm 6	84% \pm 8	<0.001

* Corrected for multiple comparisons with Bonferroni's method

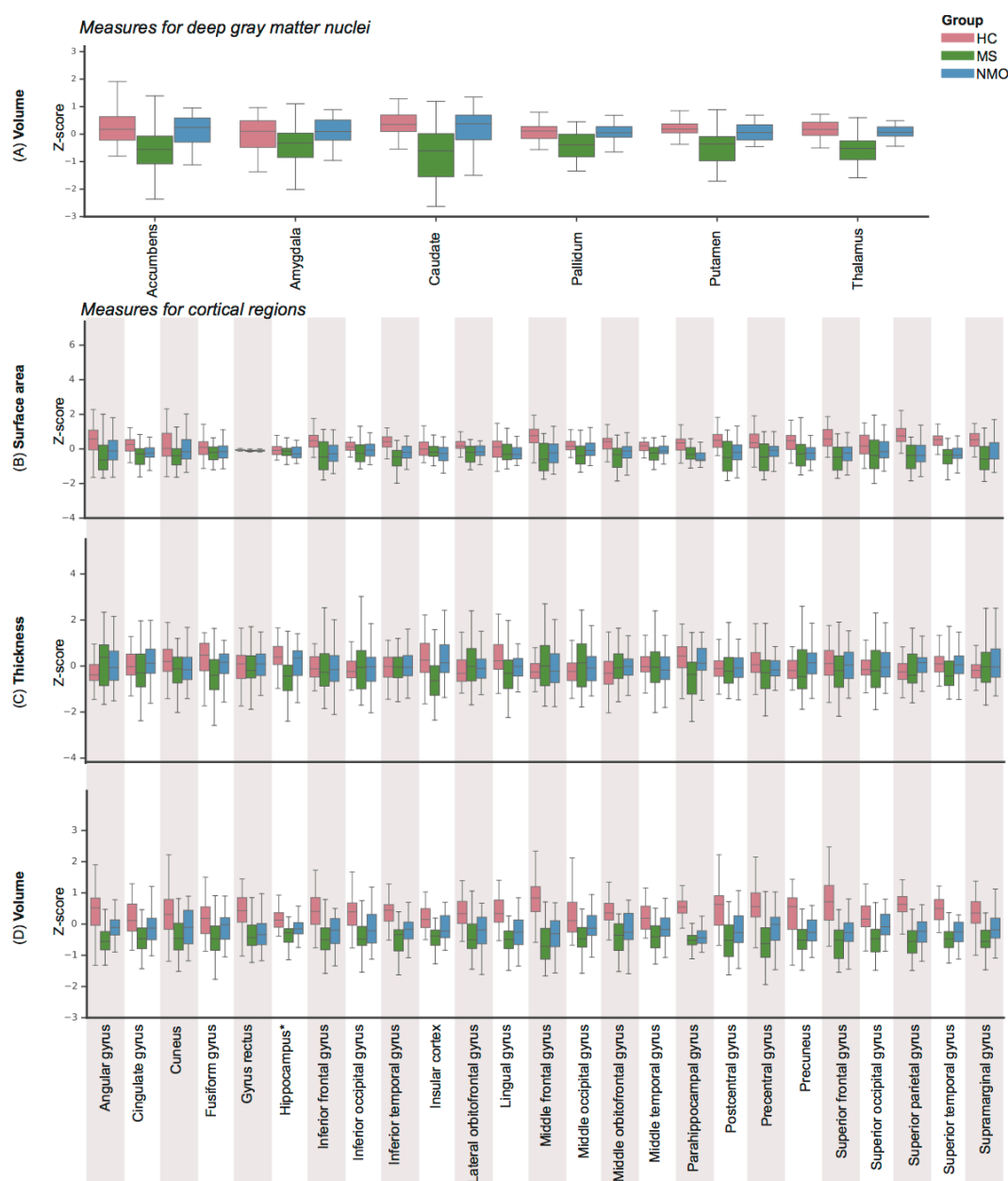
¹ multiple sclerosis; ² neuromyelitis optica; ³ healthy controls

Accuracy was defined as the percentage of patients correctly identified by the classifier as either MS or NMO. Sensitivity was defined as the number of subjects correctly identified by the classifier (MS or NMO) divided by all subjects diagnosed with each disease. Specificity is defined as the number of subjects correctly labelled as not having the disease divided by all subjects without the disease.

Variable importance. The volumes of the deep GM structures were the most important GM measures to distinguish MS from NMO, whilst the volumes of cortical regions were more important in distinguishing patients from HCs. In general, the GM volumes were more useful to the classification than the cortical thickness and surface area.

MS vs. NMO. The volumes of the deep GM (the thalami, right pallidum, and putamina) followed by the insular thickness, which showed reduction in MS compared to NMO (see **Figure 3-1** and **Figure 3-2**), were the most important GM measures to distinguish MS from NMO.

Figure 3-2. Descriptive statistics of the thickness, surface and volume.



Left and right hemisphere values are averaged for better visualization. All measures have been normalised, and the Z-scores were calculated.

When looking at individual effects of GM measures (cortical and subcortical volumes, surface area and thickness) to distinguish between NMO and MS, the combination of all measures

gave the best accuracy, which was similar to the accuracy of the classification obtained using only the deep GM volumes (**Figure 3-1**), but higher than that obtained using each measure on its own.

When I used the thalamic volume (the most discriminating GM measure) together with the white matter lesion volume (see Appendix, supplementary Figure 3-2) the average accuracy of the classification of MS vs. NMO (over 5000 permutations) was 80% (sensitivity/specificity 85%/76%, $p<0.001$).

When I trained the NMO vs MS classifier using the data from people with NMO who were AQP4 positive and tested on people with NMO who were AQP4 negative, the performance of the model was similar to that of the model that trained and tested randomly selected scans from all people with NMO regardless of serostatus (average \pm standard deviation accuracy = $77\%\pm 12$, sensitivity= 74 ± 19 , specificity= $80\%\pm 15$).

NMO vs HC. The volumes of the parahippocampal gyri and the left middle frontal gyrus were the most important variables for this classification; these regions showed smaller volumes in NMO patients as compared to HCs (see **Figure 3-2**).

MS vs HC. The volumes of the bilateral parahippocampal gyri and the right superior temporal gyrus were the most important variables for this classification (these regions showed reduced volumes in MS when compared to HCs).

Effect of centre

The variable “centre” (i.e., Tehran and Padua) was consistently found to be the least important feature for all the classifications including all variables related to volume, surface and thickness are as follows:

- MS vs NMO classification: the ratio of the importance of the most important variable (right thalamic volume) to the least important variable (centre) was: $220/0.89=247$, which means right thalamic volume was 247 times more important.
- MS vs HC classification: the ratio of the importance of the most important variable (right parahippocampal volume) vs. the least important variable (centre) was:

$338.89/0.6 = 564.81$ which means that the importance of right parahippocampal volume was 564 times more than that of centre.

- NMO vs HC classification: the ratio of the importance of the most important variable (right parahippocampal volume) vs. the least important variable (centre) was: $243/5=48.6$, which means that the importance of right parahippocampal volume was 48.6 times more than that of centre.

Classification was not influenced by the effect of centre

The models that only included centre as a predictor had the following average accuracies (\pm standard deviation) over 5000 bootstrap samples:

- NMO vs MS classification: accuracy = $47\% \pm 5.8$
- MS vs HC classification: accuracy = 57 ± 7.1
- NMO vs HC classification: accuracy = $53\% \pm 6.2$

I have reported statistical significance (p -values) against a random classifier in Table 3-2. All the three classifications remained statistically significant against a baseline classifier that had only centre as a variable (results are not shown).

3.5. Discussion

Previous imaging studies have highlighted differences in MRI characteristics between NMO and MS (Calabrese *et al.*, 2012; Duan *et al.*, 2012; Eshaghi *et al.*, 2015; Kim *et al.*, 2015), but an automatic distinction is still challenging. Here, I automatically classified patients with MS and NMO on the basis of their brain MRI scans routinely acquired with clinical protocols, using a random-forest classifier. My findings showed that GM imaging measures, such as cortical thickness, cortical surface area, and subcortical GM volumes, led to an accuracy of 74% when classifying the two patient groups, which is higher than that obtained with each GM measure on its own. When, in the *post hoc* analysis, I used both the most discriminating GM measure (the thalamic volume) and the white matter lesion volume, the accuracy of the model in classifying NMO vs MS was higher (80%) than the original model. Random forest classification accuracy was higher for distinguishing MS from HCs (92%) and NMO from HCs (88%), because the classification task is less challenging than discriminating the two diseases.

Therefore, this automatic approach may support the diagnostic process in clinical practice, for example in patients with NMOSD without AQP4. The variable “centre” did not significantly affect the classification. Moreover, the contribution of imaging features to the classification was more relevant than the differences in clinical characteristics, which may exist between patient cohorts.

I demonstrated that the volumes of the deep GM, which were lower in people with MS than NMO (**Figure 3-2**), were more important features for discriminating NMO and MS than cortical thickness and volume. My results are in line with a recent study that failed to show significant thalamic atrophy in NMO when compared to HCs (Matthews *et al.*, 2015; Finke *et al.*, 2016, p. 4), whilst another study detected a mild reduction in thalamic volume in NMO when compared to HCs (Calabrese *et al.*, 2012). The thalamic volume loss in MS may be related to secondary neurodegeneration occurring in cortical areas or may represent primary neurodegeneration (Calabrese *et al.*, 2012; Minagar *et al.*, 2013; Matthews *et al.*, 2015), whereas in NMO milder neurodegeneration secondary to axonal degeneration in the spinal cord or visual pathways is present (Liu, Fu, *et al.*, 2015).

Cortical thickness represents the number of neurons in each cortical column (Sailer *et al.*, 2003). The thickness of the insular cortices had the second highest discriminating value (after the deep GM volumes) in distinguishing MS from NMO. This region shows a predilection for atrophy, especially as patients with MS progress (Howell *et al.*, 2011). The insula is deep in the temporal lobe and may be more exposed to inflammatory mediators than other cortical areas (Howell *et al.*, 2011).

While atrophy in regions such as the thalamus and insula was only seen in MS, both demyelinating diseases seemed to affect the parahippocampal gyrus in a similar fashion (**Figure 3-2**). The parahippocampus is part of the limbic system and has extensive connections to temporal, frontal, and deep GM regions. Transection of these connections, which occurs as a result of white matter lesions, could lead to atrophy of this region via retrograde neurodegeneration in both diseases (Saji *et al.*, 2012; Calabrese, Magliozzi, *et al.*, 2015). Additionally, in MS, GM damage might be the result of either non-inflammatory or

inflammatory processes that have occurred within the grey matter itself (Geurts *et al.*, 2012; Saji *et al.*, 2012; Calabrese, Magliozzi, *et al.*, 2015). In NMO, the GM loss is more selective and depends on AQP4 expression, and the ratio of M1 to M23 proteins in astrocytes (Saji *et al.*, 2012). The parahippocampus may show a different pattern of AQP4 expression or may be connected to areas with high AQP4 expression, although the dynamic pattern of AQP4 expression in the brain remains unclear (Amiry-Moghaddam and Ottersen, 2003; Pittcock *et al.*, 2006).

Patients with NMO may have severe relapses affecting the motor and/or visual system (Wingerchuk and Weinshenker, 2014). However, GM measures of both the occipital or motor cortices did not distinguish between NMO and MS or HCs. These findings are consistent with previous voxel-based morphometry studies in NMO that failed to show detectable changes in the GM volume of these regions when compared to MS or HCs (Blanc *et al.*, 2012; Chanson *et al.*, 2013; Matthews *et al.*, 2015). Another study found mild thinning of the calcarine, precentral and postcentral cortices in NMO when compared to HCs, but no significant differences between NMO and MS (Calabrese *et al.*, 2012).

One strength of my study is that I included participants from two centres. I previously demonstrated, in a one-centre setting using scans from the same Iranian patients used in this study, that a combination of advanced MRI modalities (functional MRI and diffusion tensor imaging) distinguishes between MS and NMO with a high accuracy, using a technique called data fusion with multi-kernel learning (Eshaghi *et al.*, 2015). The limitations of this technique are that it is computationally intensive, and its application in multi-centre studies is limited. The present work, using a random-forest classification, is more robust in a multi-centre setting, and is less likely to overfit (Chen *et al.*, 2004). I included centre as a variable in my analysis and found that it had the lowest importance in all three classifications: NMO vs. MS, MS vs. HC, and NMO vs. HC. To test the effect of centre in distinguishing each pair of groups, I used the centre variable (Padua or Tehran) alone to predict the outcomes (NMO, MS or HC) and found that the resulting classifier was similar to a random classifier. Moreover, when calculating the p-values I used the baseline accuracies acquired from a classifier that used

the variable “centre” in the analysis, which did not change the results. These results confirm that my results are unlikely to be confounded by the effects of centre, although an indirect effect of centre on imaging measures cannot be ruled out.

I repeated the cross-validation 5000 times, which allowed the inclusion of all patients in both training, and test sets. This could be generalized to other centres, provided sufficient participants from new centres be included in the training sample. Moreover, for an easier application I simplified the model including only thalamic volume and white matter lesion volume; this achieved 80% accuracy in distinguishing between MS and NMO. In the future, greater accuracy may be achieved by including additional features, such as lumbar puncture results, antibody status and quantitative spinal cord measures (Lucchinetti *et al.*, 2014; Liu, Wang, *et al.*, 2015).

In this study, patients with NMO were significantly older than patients with MS, and therefore I adjusted GM measures for age. NMO has a later disease onset than MS (Wingerchuk and Weinshenker, 2014), so after I matched two groups according to disease duration, age was higher in NMO than MS. However, despite older age, which has the effect of reducing GM volume (Hedman *et al.*, 2012), patients with NMO had higher GM volume than those with MS, which suggests that age is unlikely to be a confounder in my analysis. Moreover, I have built my model using scans of patients with a secure diagnosis of MS or NMO, with a mean disease duration of 7.5 years. This is because the clinical diagnosis of MS or NMO was the gold standard against which the classification accuracy was calculated. An important question to be addressed in future studies is whether this algorithm can help to classify difficult radiological cases or to predict the outcome after the first attack. For example, in a previous study I found that machine learning correctly predicted the presence (or the absence) of clinically-definite MS in 71.4% of patients within one year of onset of a clinically isolated syndrome (Wottschel *et al.*, 2015).

This study is not without limitations. First, the NMO group, which was recruited prior to the revised criteria for the diagnosis of NMOSD (Wingerchuk *et al.*, 2015), included both AQP4 negative and positive patients (15 negative, and 50 positive). Recent work has highlighted the

possible different clinical and radiological characteristics between these two groups that may indicate distinct subtypes (Sato *et al.*, 2014). There were no notable differences in age, gender, or clinical disability between NMO without AQP4 (2 from Italy, 13 from Iran) and NMO with AQP4. However, the group of seronegative patients was small. In the future, this technique could help to identify patients with NMOSD without AQP4, since in this cohort they show similarities with the NMOSD with AQP4 cases; in particular, the algorithm correctly classified them in 77% of NMOSD-without-AQP4 cases after having learnt from only the NMOSD with AQP4 cases. Second, the role of GM lesions in distinguishing between the two diseases was not assessed, because the sequences that allow GM lesion detection are not routinely acquired in the clinical setting. Further studies will clarify these issues.

I showed that the random forest classification robustly and automatically discriminates between MS and NMO patients on the basis of their MRI scans in a two-centre setting with up to 80% accuracy. Furthermore, deep GM volumes and cortical thickness of specific key regions may give increased power to detect subtle GM features, which may facilitate the differential diagnosis between MS and NMO. In the next chapter I will build up on the results of this chapter and look at the longitudinal evolution of brain volumes in MS patients of different phenotypes.

4. Deep grey matter volume loss drives disability worsening in multiple sclerosis

4.1. Summary

Grey matter (GM) atrophy occurs in all multiple sclerosis (MS) phenotypes. I investigated whether there is a spatiotemporal pattern of GM atrophy that is associated with faster disability accumulation in MS. I analysed 3,604 brain high-resolution T1-weighted MRI scans from 1,417 participants: 1,214 MS patients (253 clinically-isolated syndrome [CIS], 708 relapsing-remitting[RRMS], 128 secondary-progressive[SPMS], 125 primary-progressive[PPMS]), over an average follow-up of 2.41 years (standard deviation[SD]=1.97), and 203 healthy controls (HCs) [average follow-up=1.83 year, SD=1.77], attending 7 European centres. Disability was assessed with the Expanded-Disability Status Scale (EDSS). I obtained volumes of the deep GM (DGM), temporal, frontal, parietal, occipital and cerebellar GM, brainstem and cerebral white matter. Hierarchical mixed-models assessed annual percentage rate of regional tissue loss and identified regional volumes associated with time-to-EDSS progression. SPMS showed the lowest baseline volumes of cortical GM and DGM. Of all baseline regional volumes, only that of the DGM predicted time-to-EDSS progression (hazard ratio=0.73, 95% CIs 0.65, 0.82; $p<0.001$): for every standard deviation decrease in baseline DGM volume, the risk of presenting a shorter time to EDSS worsening during follow-up increased by 27%. Of all longitudinal measures, DGM showed the fastest annual rate of atrophy, which was faster in SPMS (-1.45%), PPMS (-1.66%), and RRMS (-1.34%) than CIS (-0.88%) and HCs (-0.94%)[$p<0.01$]. The rate of temporal GM atrophy in SPMS (-1.21%) was significantly faster than RRMS (-0.76%), CIS (-0.75%), and HCs (-0.51%). Similarly, the rate of parietal GM atrophy in SPMS (-1.24%) was faster than CIS (-0.63%) and HCs (-0.23%) (all p values <0.05). Only the atrophy rate in DGM in patients was significantly associated with disability accumulation (beta=0.04, $p<0.001$). This large multi-centre and longitudinal study shows that DGM volume loss drives disability accumulation in MS, and that temporal cortical GM shows accelerated atrophy in SPMS than RRMS. The difference in regional GM atrophy development

between phenotypes needs to be taken into account when evaluating treatment effect of therapeutic interventions.

4.2. Introduction

As explained in Chapter 1 the clinical course of multiple sclerosis (MS) is heterogeneous. Some patients experience relapses with recovery (relapsing-remitting [RR] MS), while others develop progressive disability either from the onset (primary-progressive [PP] MS), or after a period of relapses (secondary-progressive [SP] MS). RRMS patients account for approximately 90% of cases at onset (Browne *et al.*, 2014), whose majority later progress to SPMS. The pathogenic mechanisms driving accrual of disability are beginning to be elucidate (Mahad *et al.*, 2015): neurodegeneration plays a crucial role in determining accrual of disability over time (Geurts *et al.*, 2012).

As shown in a cross-sectional analysis in the previous chapter, patients with RRMS have lower volumes in several brain regions that can distinguish them from neuromyelitis optica and HCs. Over time, brain volume declines more rapidly in MS patients when compared with age-matched healthy controls (HCs) (Bermel and Bakshi, 2006; De Stefano *et al.*, 2010, 2015; Geurts *et al.*, 2012). Across MS phenotypes, SPMS shows the fastest annual rate of brain atrophy, which is estimated to be 0.6% (compared to about 0.2% in age-matched HCs) (De Stefano *et al.*, 2010). The role of brain atrophy in monitoring response to treatments in MS is evolving: whole brain atrophy has been recently used as the primary outcome measure in phase 2 clinical trials in SPMS (Chataway *et al.*, 2014).

Whole brain atrophy is mainly driven by the neuroaxonal loss in GM (see Chapter 1 for more details). GM volume loss is associated with long-term disability (Fisher *et al.*, 2008; Fisniku *et al.*, 2008), and explains physical disability better than white matter (Fisher *et al.*, 2008; Roosendaal *et al.*, 2011) and whole brain atrophy (De Stefano *et al.*, 2010). Some GM regions, such as the cingulate cortex and thalamus, are affected by volume loss more extensively than others (Eshaghi *et al.*, 2014; Steenwijk *et al.*, 2016), and the extent of their volume loss correlates with disability, and cognitive impairment (Schoonheim *et al.*, 2015). Regional predilection for atrophy is not unique to MS; for example, hippocampal atrophy is more pronounced than the whole brain atrophy in the early phase of Alzheimer's disease (Henneman *et al.*, 2009). Although cross-sectional studies have previously shown patterns of

regional atrophy in different types of MS (Ceccarelli *et al.*, 2008; Steenwijk *et al.*, 2016), studies on longitudinal evolution of atrophy in different structures across MS phenotypes are scarce. The overarching goal of this chapter was to investigate whether there is a spatiotemporal pattern of GM atrophy that is associated with faster disability accumulation in MS. In a large multi-centre cohort, which included all MS phenotypes and HCs, I tested the following hypotheses: (i) some GM regions show faster atrophy rate than others and their rate may differ between MS phenotypes; (ii) smaller baseline volumes of brain structures, reflecting a more extensive neurodegeneration, predict disability accrual; (iii) the rate of regional volume loss is associated with the rate of disability accumulation.

4.3. Methods

4.3.1. Participants

In this retrospective study, I collected data from 7 European MS centres (MAGNIMS: www.magnims.eu) from 1,424 participants who have been studied between 1996 and 2016; I included participants who fulfilled the following criteria: (1) a diagnosis of MS according to 2010 McDonald Criteria (Polman *et al.*, 2011) or a clinically isolated syndrome (CIS, see **Chapter 1** for the diagnosis criteria and related references); (2) healthy controls (HCs) without history of neurological or psychiatric disorders; (3) at least two-MRI scans acquired with a minimal interval of 6 months with identical protocol, including high-resolution T1-weighted MRI (allowing regional grey and white matter segmentation), and T2-weighted/Fluid Attenuated Inversion Recovery (FLAIR), sequences. Patients were scored on Expanded Disability Status Scale (EDSS). To increase the number of HCs scans, which were provided by four centres, I collected data from age-matched HCs from the Parkinson's Progression Marker's Initiative (<http://www.ppmi-info.org/data>). Part of these data have previously been used in other publications (Roosendaal *et al.*, 2011; Eshaghi *et al.*, 2014; Cawley *et al.*, 2015; Brown *et al.*, 2017).

MRI scans were taken under consent obtained from each subject independently in each centre. The final protocol for this study was reviewed and approved by the European MAGNIMS collaboration for analysis of pseudo-anonymised scans.

4.3.2. Image acquisition

I included scans from 13 different MRI protocols; all centres except one provided 3D-T1 weighted scans (**Appendix: Supplementary Table 4-1** and **Supplementary Table 4-2** show the MRI protocols).

4.3.3. Image analysis

I performed image analysis as follows:

1) Bias field correction

I used N4 bias field correction to correct for field inhomogeneity in T1-weighted scans using ANTs v2.10 (Tustison *et al.*, 2010).

2) Lesion filling

Lesion masks were manually delineated on PD/T2-weighted images by different raters at each centre semi-automatically, except for three centres that used the same automatic lesion segmentation with LST toolbox (version 2.0.15) (Schmidt *et al.*, 2012). I calculated linear transformation matrices to register T2-weighted/FLAIR with the T1-weighted scan using FSL-FLIRT v5.0 (Jenkinson and Smith, 2001). Then I applied these matrices to lesion masks to transfer them into the accompanying T1-weighted subject-space. I used the FSL lesion filling method which uses a white matter mask calculated with FSL-FAST (Zhang *et al.*, 2001, p.) to fill T1-weighted hypo-intensities within normal-appearing whiter matter, so to reduce segmentation errors, as previously done (Amato *et al.*, 2012; Battaglini *et al.*, 2012; Popescu *et al.*, 2013).

3) Symmetric within-subject registration

To avoid asymmetric registration and interpolation of longitudinal scans (e.g., toward the baseline scan), I constructed an unbiased subject-specific template that has “equal distance” from each time point using FreeSurfer version 5.3 (Reuter *et al.*, 2010, 2012; Reuter and Fischl, 2011). I linearly transformed T1-weighted images to this symmetric space with the unbiased transformation matrix for each time point and used cubic B-spline interpolation to reduce interpolation artefacts. I manually checked the alignment of scans in the symmetric space.

4) Tissue segmentation

Next, in the symmetric space, I segmented T1-weighted scans into the GM, white matter and cerebrospinal fluid with the Geodesic Information Flow (GIF) software (part of NiftySeg, <http://cmictig.cs.ucl.ac.uk/niftyweb/program.php?p=GIF>) (Cardoso *et al.*, 2015), and parcellated each hemisphere into regions of interest according to the Neuromorphometric atlas (Klein and Tourville, 2012). GIF uses an atlas propagation and label fusion strategy to calculate the voxel probabilities of GM, white matter and CSF (Cardoso *et al.*, 2015); this method has been previously used in MS and other neurodegenerative disorders (Bocchetta *et al.*, 2016; Pardini *et al.*, 2016). The template library had 95 MRI brain scans (HCs and

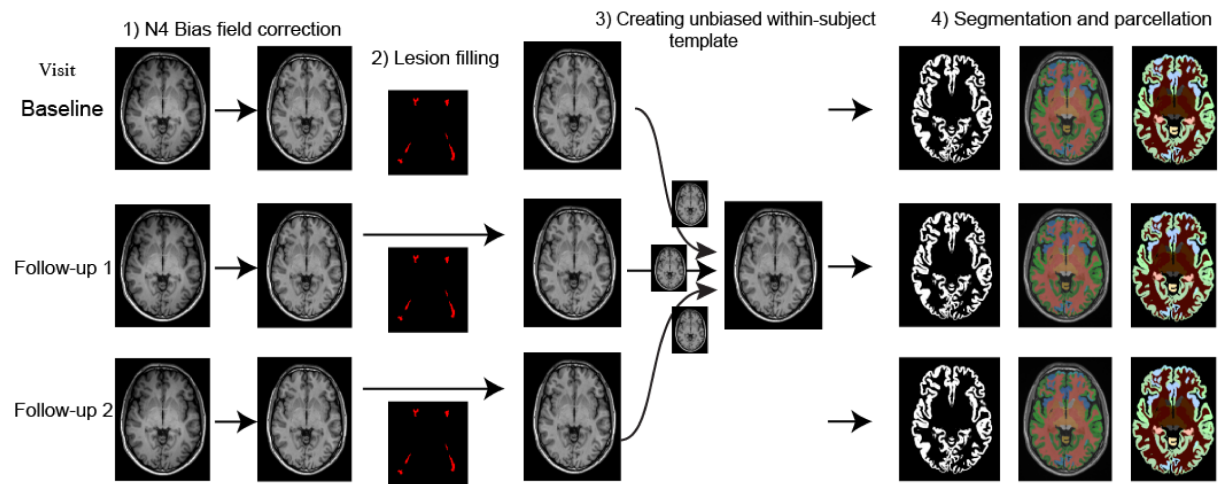
patients with Alzheimer's disease) with neuroanatomic labels (<http://www.neuromorphometrics.com/>). This atlas, which is similar to Mindboggle atlas, was developed to improve the consistency and clarity of Desikan-Killiany protocol (Klein and Tourville, 2012).

To calculate brain masks and exclude segmentation errors outside of the brain I used STEPS (Similarity and Truth Estimation for Propagated Segmentations, <http://cmictig.cs.ucl.ac.uk/niftyweb/program.php?p=BRAIN-STEPS>) based on a template library of 682 hand-drawn brain masks (Jorge Cardoso *et al.*, 2013; Prados *et al.*, 2015). These maps were applied to each time point separately.

5) *Regional volume calculation*

I visually assessed the segmentations to assure the quality for statistical analysis. To calculate regional volumes, I summed the probability of the segmented tissue voxels (GM or white matter) in each parcellated region and multiplied the sum with the voxel volume. I averaged values between left and right hemispheres. Next, I summarised the regional volumes according to Neuromorphometrics protocol by summing the volume of GM regions in the temporal, parietal, occipital, frontal lobes, cerebellum and deep GM (DGM) [thalamus, putamen, globus pallidus, caudate, and amygdala]. I also obtained the volume of the brainstem and the cerebral white matter. **Figure 4-1** shows the image analysis pipeline.

Figure 4-1. Imaging analysis pipeline.



An unbiased symmetric image registration approach was used to calculate atrophy.

4.3.4. Statistical analysis

Brain volumes at baseline and rates of volume changes over time

To investigate baseline volumes (intercept) and rates (slopes) of volume change by subject group and region, I used linear mixed-effects models with the volume at a given time as the response variable, and time and interactions with time as fixed-effect covariates (Bernal-Rusiel *et al.*, 2013). This model estimates adjusted rate while allowing for nested correlation structures, such as time of visit within subject within scanner, by incorporating, in this example, subject and scanner random intercepts, and a random slope on time. The interaction terms with time (e.g., subject group X time), allows the estimation of rate differences across the interacting variable, in this example subject groups or clinical phenotypes. Including another interaction with time, such as gender X time, adjusts the rate for gender. In addition to time, the fixed-effect covariates were: scanner magnetic field, subject group, gender, age at baseline, total intracranial volume (sum of the volumes of GM, WM and CSF) at baseline; and the interactions of each of these with time. Disease duration was too highly correlated with age at baseline to give a reliable estimation, and was omitted from the final models. To estimate the percentage changes per unit (year) increase in time, I log-transformed the volume (Vittinghoff *et al.*, 2006). I adjusted time to zero for those visits in which a patient converted from one phenotype to another (e.g., CIS to RRMS). I performed *post-hoc* analyses to identify

specific GM regions within the cerebral lobes and among the DGM nuclei that showed significant differences between MS phenotypes, as well as the default-mode network regions (Zhang and Raichle, 2010).

To investigate whether there is an association between the rate of loss in specific regions and MS phenotypes, 3-way interactions were used, for example, clinical phenotype \times region \times time. I used R (version 3.2.2) and the NLME package (Pinheiro and Bates, 2009; R Core Team, 2014).

For each model, I visually checked the heteroscedasticity (which is the unequal variance of a variable across the range of values of a second variable that predicts it) per group by plotting residuals against the fitted values.

I corrected for multiple comparisons accounting for the number of all the tests performed with the false-discovery rate method (Benjamini and Hochberg, 1995).

Effect of MRI protocols on imaging measures

To assess the effect of the MRI protocol on MRI measures (I took into account the protocols rather than the centres because some centres acquired more than one protocol with more than one scanner) I included it as a fixed-effect variable in a separate mixed-effect model, and calculated the average effect sizes for MRI protocols and MS phenotypes (i.e., disease effects) while fixing other variables.

Assessing associations between brain tissue volumes and disability accrual

For easier interpretation of clinical and imaging measures, I standardised volumes by subtracting the mean and dividing by the standard deviation (Z-score). I analysed CIS and relapse-onset patients together, because some patients had converted from CIS to RRMS, or from RRMS to SPMS. This allowed us to take advantage of a longer follow-up period. With similar mixed-effects models I investigated the following three questions: (1) Are the baseline volumes of the DGM, the temporal, frontal, parietal, occipital and cerebellar GM, brainstem, and white matter, and white matter lesion load associated with EDSS at baseline? (2) Are changes in all these regional volumes and white matter lesion load associated with EDSS changes over time? (3) Do baseline volumes of all these regions and white matter lesion at

baseline predict time-to-EDSS progression (event=EDSS progression) during follow-up? The EDSS-progression event was defined as 1.5 increase in EDSS, if the baseline EDSS was 0; one-point increase if EDSS was less than or equal to 6; and 0.5 increase if EDSS was more than 6 (Healy *et al.*, 2013). I used a Cox-regression model to explore whether baseline volumes of these structures predicted time to event. I performed a *post-hoc* analysis using all GM regions to determine the most important predictors of time-to-EDSS-progression (as defined above) and confirm that the results of the DGM were not affected by the bias of merging a higher number of cortical regions into the main lobes. I performed FDR correction to adjust for multiple comparisons (Benjamini and Hochberg, 1995).

Additional analyses: software reliability and effects of disease-modifying treatments

I carried out additional analyses to assess the reliability of brain volumes estimated with GIF software, FSL-FIRST, and SPM12, and effects of treatments on atrophy measures.

To confirm whether the results obtained with GIF were reproduced by using more established pipelines, I performed additional analyses to obtain volumes of DGM structures and cortex using FSL (version 5.0.9) FIRST, and SPM12 (Ashburner and Friston, 2005), respectively. I randomly sampled half of subjects irrespective of centre or MS phenotype. I fed T1-weighted output images from step 3 of the pipeline into FSL-FIRST and SPM12 software. Details of segmentation algorithms are explained elsewhere (Ashburner and Friston, 2005; Patenaude *et al.*, 2011). I excluded subcortical structures from SPM segmentation by applying the inverse of the FSL-FIRST mask.

I calculated total DGM volumes from GIF and FSL-FIRST, and total cerebellar and cortical GM volumes from GIF and SPM12 at each visit. I separately assessed between-software reliability for DGM and cortical GM volumes, while adjusting for repeated measures and the effects of centre. I constructed two multi-level mixed-effects models in which random effects were “software” (GIF, and SPM12 for one and GIF or FSL-FIRST for the other) nested in visits, which was nested in subjects, which was in turn nested in centre. I followed the procedure for longitudinal MRI volumetric reliability explained in Cannon *et al.* (Cannon *et al.*, 2014) to calculate intra-class correlation coefficients (ICC). I defined ICC=

$\frac{\delta^2_{\text{centre}} + \delta^2_{\text{subject}} + \delta^2_{\text{visit}}}{\delta^2_{\text{centre}} + \delta^2_{\text{subject}} + \delta^2_{\text{visit}} + \delta^2_{\text{software}} + \delta^2_{\text{residual}}}$, which is the fraction of variance attributable to centre,

subject, and visits. This reflects longitudinal reliability between software for the same centre, the same subject at the same visit, in a spectrum in which an ICC of 1 indicates perfect agreement between software, and 0 indicates no agreement. I used NLME package and R version 3.2.2.

I also performed area under the curve (AUC) analysis to examine the prognostic accuracy of adjusted DGM volumes at individual level.

Effect of disease-modifying drugs

For each MS phenotype, I constructed a mixed-effects model as explained above, and looked at the differences between patients who were receiving disease-modifying treatments and those who were not.

Can a cut-off value for deep grey matter volume be used at the individual level to prognosticate patients?

I performed the area-under-the-curve (AUC) analysis to examine whether different cut-off values of adjusted baseline DGM volumes could be used in individuals to predict future EDSS-progression (as defined in the manuscript) at 6 months, 1 year, 2 year or 5 years. This method, which is explained in detail elsewhere (Patenaude *et al.*, 2011), calculates time-dependent sensitivity, and specificity to characterise prognostic accuracy of a biomarker (DGM volume).

4.4. Results

The MRI scans of 1,417 subjects were analysed (scans of three subjects were excluded due to significant motion artefacts on visual inspection and four due to registration issues because of missing MRI header information); 1,214 patients (253 had CIS, 708 had RRMS, 128 had SPMS, and 125 had PPMS), and 203 were HCs. In total, I analysed 3,604 T1-weighted MRI. Average number of scans per subject was 2.54 (SD=1.04), with an average follow up of 2.41years (SD=1.97) for patients, and 1.83 (SD=1.77) years for HCs (see **Table 1** for follow-up information per group). The total numbers of participants with 3 or more visits for each group were: 90 HCs, 48 CIS, 334 RRMS, 39 SPMS, and 58 PPMS. A total of 96 patients with CIS (38%) converted to RRMS, and 28 patients with RRMS (4%) converted to SPMS during the follow-up.

There was a significant difference in gender ratio between groups ($p<0.001$, see Table 4-1 for gender ratios). Patients with progressive MS (SPMS and PPMS) had significantly greater disability than patients with RRMS and CIS (Mann-Whitney tests, $p<0.001$, see Table 4-1), and were older than RRMS ($p<0.001$, average difference=10.7 years), CIS ($p<0.01$, average difference=15.6 years) and HCs ($p<0.01$, average difference=10 years). Age was similar between patients with RRMS and HCs. Patients with CIS were younger than HCs ($p<0.01$, average difference=4.9 years). Patients with CIS had the lowest T2-weighted lesion load, and patients with SPMS had the highest T2-weighted lesion load. About half of patients with RRMS were on disease-modifying treatments (see **Table 4-1**).

Table 4-1 Baseline characteristics of participants

Group	Healthy controls	CIS	RRMS	SPMS	PPMS
Total number (number of females)	203 (112)	253 (171)	708 (473)	128 (75)	125 (55)
Average follow up in years (range)	1.83 (0.5-7.8)	1.46 (0.5-13)	2.72 (0-13)	2.06 (0-5.5)	2.85 (0.5-6)
Average age (\pm SD)	38.7 \pm 10.5	33 \pm 8	38.2 \pm 9.8	48.2 \pm 9.8	48.5 \pm 10.1
Average disease duration (\pm SD)	—	0.4 \pm 1.4	6.7 \pm 7.3	15.6 \pm 9.9	6.8 \pm 5.9
Median EDSS (range)	—	1 (0-4.5)	2 (0-7)	6 (2.5-9)	5 (2-8)
Median T2 lesion load (ml) (1st-3rd quartiles)	—	2.97 (1.01-5.04)	5.05 (2.05-11.79)	11.04 (3.18- 23.14)	9.38 (2.69- 22.02)
% (number) of patients on DMTs	—	20% (52)	49% (345)	41% (52)	6% (8)

Table legend: Abbreviations: SD, standard deviation; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary-progressive multiple sclerosis; PPMS, primary-progressive multiple sclerosis; ml, millilitre; EDSS, expanded-disability status scale; DMTs: disease-modifying treatment.

4.4.1. Brain atrophy at baseline in MS and rates of volume changes over time

At baseline, all clinical phenotypes (CIS, RRMS, SPMS, and PPMS) had significantly smaller cortical GM and DGM volumes than HCs. SPMS showed the lowest cortical GM and DGM volumes, followed by PPMS, RRMS, CIS. All clinical phenotypes, but not CIS, had significantly reduced whole brain and white matter volumes when compared to HCs (see **Figure 2A**).

The fastest regional decline in tissue volume over time was seen in the DGM in all clinical phenotypes (PPMS: -1.66% per year, SPMS: -1.45%, RRMS: -1.34%, CIS: -0.88%, $p < 0.01$) and in HCs (-0.94%). The rate of atrophy in the DGM was greater in RRMS, SPMS and PPMS than CIS and HCs (all p values < 0.01) (**Figure 2B** and **Supplementary Tables 4-3 and 4-4** in the Appendix of thesis), but did not differ between RRMS, SPMS and PPMS. The rate of volume loss in the DGM in all MS patients together was significantly higher than that in the cortical and cerebellar GM and brainstem (although the rate of volume loss over time in these areas was still significant) (all p values < 0.05).

The volume loss of the whole cortical GM was faster in SPMS (-1.11% per year), PPMS (-0.79%), RRMS (-0.67%), than HCs (-0.34%) (all p values < 0.05). Among the cortical regions, the temporal lobe GM showed a faster volume loss in SPMS (-1.21%) than RRMS (-0.77%) and CIS (-0.75%) (all p values < 0.05) (**Figure 4-2B** and **Supplementary Tables 4-3 and 4-4** in Appendix of thesis). Similarly, the parietal GM showed a faster volume loss in SPMS (-1.24%) than CIS (-0.63%) ($p < 0.05$) (**Figure 4-2B** and **Supplementary Tables 4-3 and 4-4** in the Appendix). No differences in rates of volume loss were seen in the frontal and occipital GM between clinical phenotypes. Overall, all the cortical GM regions, except for the occipital cortex, showed a faster rate of atrophy in MS than HCs (**Figure 4-2B** and **Supplementary Table 4-4** in the Appendix).

The white matter did not show a significant rate of volume loss in HCs or any of the clinical phenotypes.

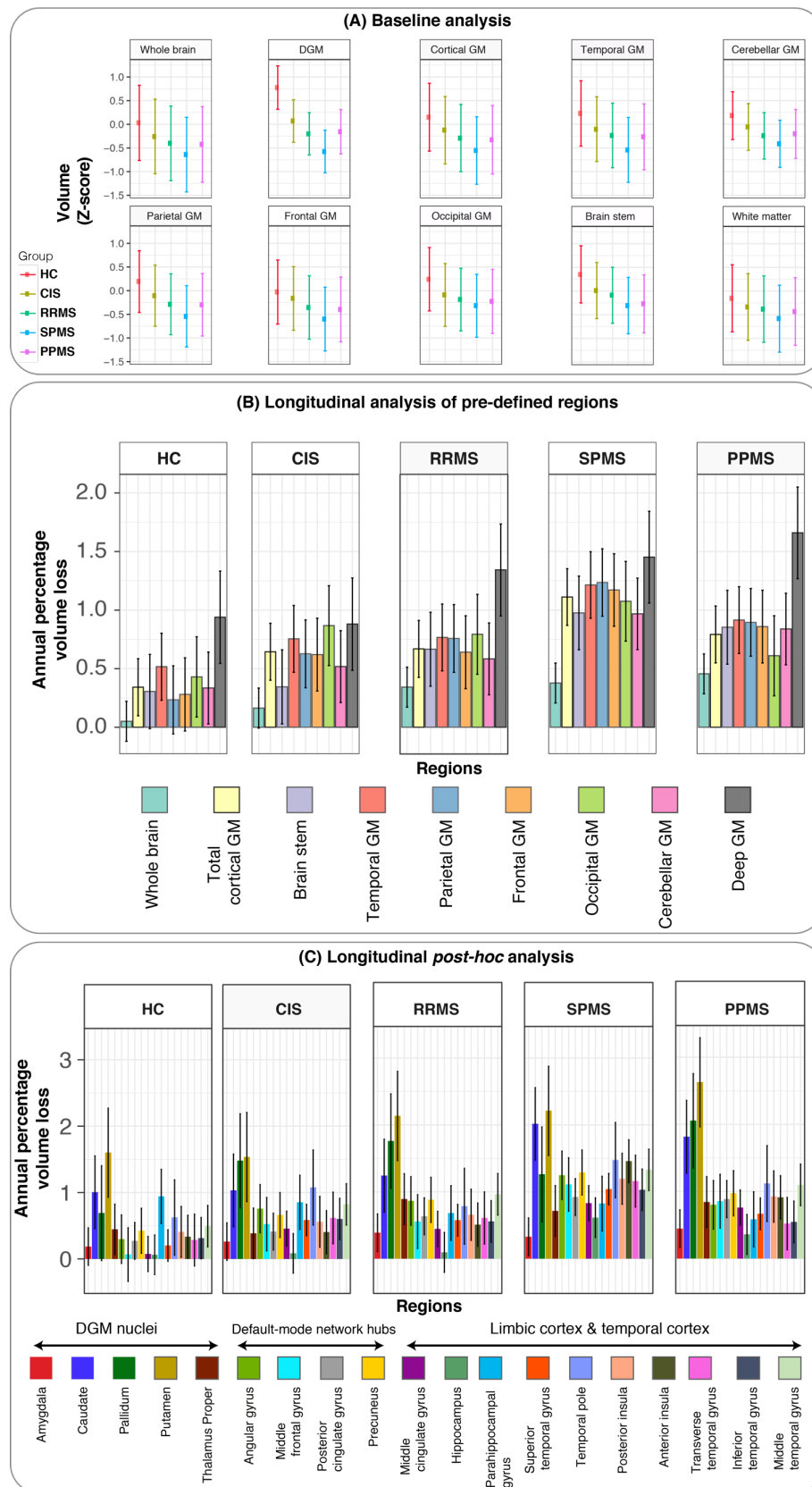
There was no heteroscedasticity in the plots of residuals against fitted values.

In the *post-hoc* analyses when looking at regions and clinical phenotypes I found that among the DGM nuclei, the putamen showed the fastest volume loss in PPMS (-2.6%). Within the

temporal lobe GM, the fastest volume loss was seen in the temporal pole (-1.47%) and posterior insula in SPMS (-1.19%). When looking at the parietal lobe GM, the precuneus showed the fastest atrophy rates in SPMS (-1.28%) (**Figure 2C**). Whilst the fastest rate of atrophy was seen in DGM in SPMS, the temporal lobe GM showed the highest difference between SPMS and HCs (see **Figure 2C**).

There was no significant effect of gender on rates of atrophy. There was no significant association between GM volumes and T2-weighted (or FLAIR) lesion load.

Figure 4-2. Baseline volumes, and annual percentage loss of brain regions in clinical phenotypes and healthy controls.



Adjusted baseline values for HCs, CIS, RRMS, SPMS, and PPMS are shown in (A), where the adjusted mean is shown as a point, and error bars show the 95% confidence-interval. Adjusted P-values of pairwise comparisons between groups are shown in Supplementary Table 4. Longitudinal analyses are shown in (B) and (C). Bar charts of the adjusted annual percentage of loss are shown in (B) for the predefined regions. The height of each bar chart is the average estimate of the annual percentage loss from the mixed-effects model for each group. The error bars represent 95% confidence interval of these estimates. Adjusted P-values for pairwise comparison between regions across clinical phenotypes and HCs are shown in Supplementary Table 4. White matter volumes are not shown in (B, and C) because they did not show a significant change over time in any clinical phenotype. Post-hoc analyses of annual percentage loss are shown in (C) where DGM nuclei, temporal, limbic and default mode network regions were selected. Similar to (B) the adjusted average annual percentage volume loss for these regions is the height of each bar-chart and error bars represent 95% confidence intervals.

Baseline values (A) and rates (B, and C) were adjusted in a single mixed-effects hierarchical model including age, gender, total intracranial volume at baseline, scanner magnetic field, and their interactions with time as the fixed-effects. Centre, subject and visits were nested (hierarchical) random-effects.

Abbreviations: HC, healthy controls; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary-progressive multiple sclerosis; PPMS, primary-progressive multiple sclerosis.

4.4.2. Regions showing the highest rate of loss

When I compared the rate of volume loss across different regions in all patients (CIS, RRMS, SPMS, and PPMS) together, the fastest decline (or lowest slope) was seen in the DGM (**Supplementary Tables 4-3 and 4-4** in the Appendix). The rate of loss in the cortical GM

regions was similar between lobes and to that of the cerebellum. The slowest rate of loss was seen in the brainstem.

4.4.3. Spatiotemporal pattern of GM volume loss in clinical phenotypes

Although SPMS showed the lowest baseline volumes of cortical GM and DGM, and the rate of the DGM volume loss was faster in SPMS, PPMS and RRMS than CIS and HCs, there was no significant association between the rate of loss in specific regions and clinical phenotypes, which suggests that all clinical phenotypes share a similar spatiotemporal pattern of GM loss.

4.4.4. Effect of MRI protocols on imaging measures

The average effects of MS phenotypes on brain volumes at baseline were higher than the protocol effect on the brain volumes (protocol effects: whole brain = 4.3%, cortical GM = 5.1%, DGM = 8.5%, disease effects: whole brain = 4.8%, cortical GM = 5.2%, DGM = 13.7%). The average effects of MS phenotypes were higher than the effects of protocol on the rates of atrophy of the cortical GM and DGM (protocol effects: cortical GM = 0.14%, DGM = 0.21%, disease effects: cortical GM = 0.57%, DGM = 0.53%), but not those of the whole brain (protocol effect = 0.51%, and disease effect = 0.38%).

4.4.5. Association between EDSS and GM loss

In all clinical phenotypes combined, lower DGM and cortical GM volumes at baseline were associated with higher disability, as measured by the EDSS (DGM $\beta = -0.71$, $p < 0.0001$; cortical GM $\beta = -0.22$, $p < 0.0001$). Under the assumption of a linear relationship between EDSS and GM volume, this suggests that for every Z-score decrease in the DGM and cortical volume at baseline, the baseline EDSS increased on average by 0.7 and 0.22, respectively.

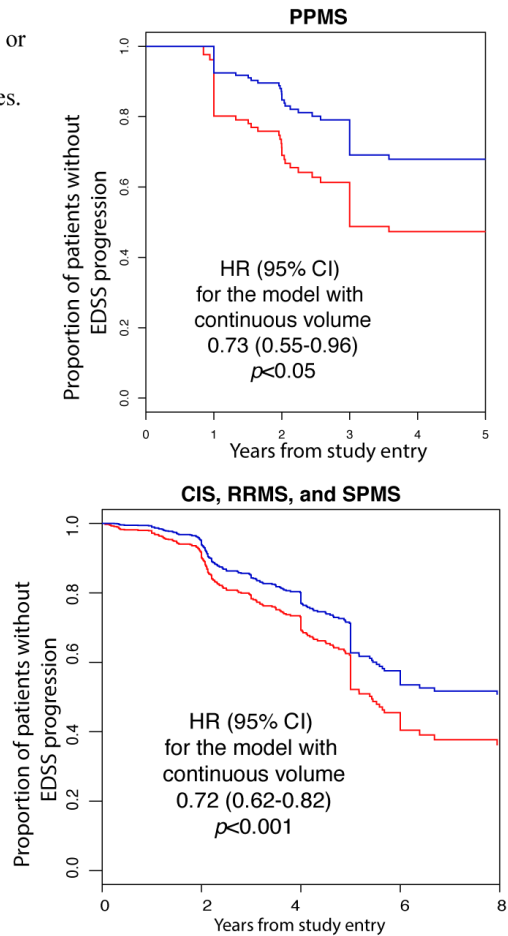
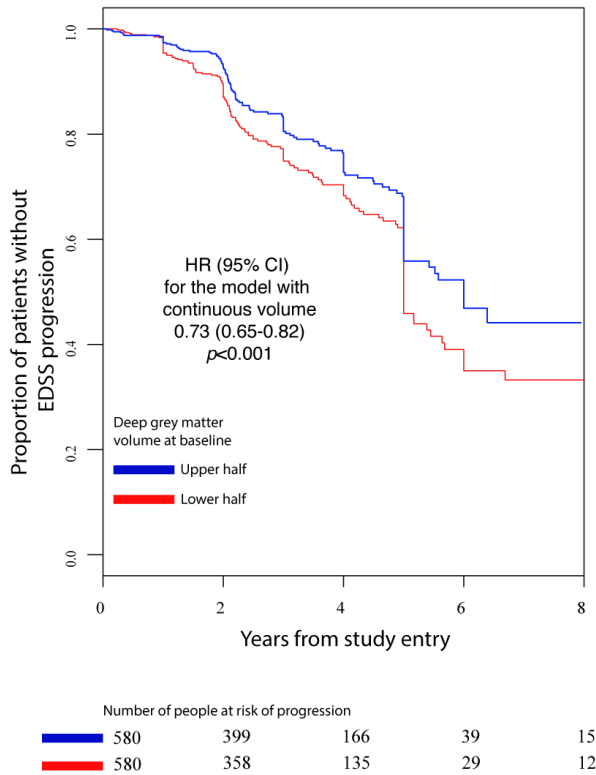
There was a significant progression of EDSS in both relapse-onset and PPMS patients, which on average increased by 0.07 and 0.2 per year, respectively. When I examined associations between the rate of EDSS changes and rate of changes in the volumes of cortical GM regions, cerebellar GM and DGM over time, only the rate of loss in the DGM was associated with disability accumulation ($\beta = -0.04$, 95% CI: -0.02, -0.06, $p = 0.006$). Under the assumption of a linear relationship between EDSS and rate of GM volume loss over time, this suggests that

every standard deviation (Z-score) loss in the rate of DGM volume corresponded to an annual EDSS gain of 0.04.

The percentage of patients who had EDSS progression during follow-up (or who experienced the “event”) was 26%. When I looked at baseline predictors of disability accumulation, without any longitudinal imaging measure in the model, only the DGM predicted future EDSS progression. The hazard ratio [95% CI, *p*-value] for time-to-EDSS progression was 0.73 [95% CI 0.65, 0.82, *p*<0.0001], which suggests that for every standard deviation (Z-score) decrease in the DGM volume at baseline the risk of presenting a shorter time to EDSS worsening during the follow-up increased by 27% [95% CI: 18-35%]. The hazard ratio remained similar when I analysed relapse-onset and PPMS patients separately (0.72 and 0.73 respectively). **Figure 4-3** illustrates the survival-curve for these analyses.

Figure 4-3. DGM volume predicts future progression of EDSS.

(A) **Baseline** DGM volume, but not baseline lobar cortical grey matter or whol brain volumes, can predict **future** EDSS progression. Predictive value of DGM volume is independent of clinical phenotypes.



Survival curves for time to event (sustained EDSS progression, see methods for definition) in CIS, relapse-onset and PPMS. I have analysed CIS and relapse-onset patient together, because a proportion of patients convert from CIS to RRMS, or from RRMS to SPMS during the course of study. Hazard-ratios for models with continuous outcome variables (regional volumes) are reported.

In the *post-hoc* analyses, baseline thalamic volume had the highest predictive value of EDSS-progression during follow-up in both PPMS and the relapse-onset groups, by increasing the risk to a shorter time to EDSS worsening of 37% in relapse-onset MS and 40% in PPMS (**Figure 4-4B** and **C**). In this analysis, the predictive value of the thalamus was followed by that of the hippocampus and angular gyrus in relapse-onset MS (**Figure 4-4B**), and by that of the putamen, posterior insula and temporal pole in PPMS (**Figure 4-4C**).

and (C). The y-axes show the risk of progression for each Z-score loss in the volume of the corresponding brain region on x-axes. For example, for every Z-score loss of the thalamus volume at baseline, the risk of EDSS worsening during follow-up increased by 37% for the CIS, RRMS, SPMS group, and 40% for PPMS. Colour maps code the importance of baseline volumes of the regions to predict EDSS-worsening (or EDSS-progression) during follow-up. The absolute values of coefficients for ventricular volumes are shown in (B), because they have an effect in the opposite direction of other structures. Error-bars indicate the 95% confidence intervals.

DGM volume measurement reliability

The random sample included 1065 T1-weighted scans which belonged to 700 subjects of whom 116 were HCs, 151 were CIS, 328 were RRMS, 45 were SPMS and 60 were PPMS at the study entry. One visit of a subject was excluded from the analysis after failure in FIRST segmentation. GIF volumes showed excellent agreement with that of FSL-FIRST (ICC=0.87).

Cortical and cerebellar GM volume measurement reliability

GIF volumes showed excellent agreement with that of SPM12 volumes with an ICC of 0.88.

No effect of treatment

Treatment information is shown in Table 1. Among patients with available treatment information, 90% were receiving first-line injectable drugs (interferon or glatiramer acetate). People with CIS who were receiving disease-modifying drugs had smaller volumes in the cortical GM and deep grey than patients with CIS who were not ($p<0.05$) at the study entry. There was no significant difference in the baseline volumes in other MS phenotypes between patients who were on treatment and those who were not. The rates of atrophy in the whole brain, cortical GM, and DGM were similar ($p>0.05$) across all the MS phenotypes (including CIS) between patients who were on disease-modifying treatments and those who were not.

Predictive accuracy of DGM volume at the individual level is not significant

AUCs for different lengths of follow-up (6-month AUC= 0.52, 1-year AUC = 0.48, 2-year AUC = 0.47, and 5-year AUC = 0.53) were not significantly different from a random classifier.

4.5. Discussion

In this chapter I built up on the cross-sectional evidence from previous chapter and extended the analysis of grey matter volumes to a longitudinal setting . In this chapter in a large multicentre study, I have shown that volume loss in DGM over time was faster than that seen in other brain regions across all clinical phenotypes, and DGM volume loss was the only GM region associated with disability accumulation. Additionally, I found that the smaller DGM volume at baseline was associated with increased risk of shorter time to EDSS progression, in agreement with previous studies that showed smaller DGM volume associated with higher disability (Rocca *et al.*, 2010; Schoonheim *et al.*, 2015). Interestingly, I found that atrophy rates of the GM of cortical lobes were the fastest in SPMS, and were faster in the temporal lobe in SPMS in comparison with RRMS and CIS and in the parietal lobe in SPMS in comparison with CIS. However, no significant association between cortical regions and disability progression was detected. Overall, my findings suggest that the development of DGM atrophy may drive disability accumulation irrespective of clinical phenotypes, thereby becoming a useful outcome measure in neuroprotective clinical trials. Although the spatiotemporal pattern of atrophy remains similar across MS phenotypes, some cortical regions show accelerated atrophy in SPMS than RRMS and/or CIS. Moreover, this chapter extends the results of **Chapter 3** to a longitudinal setting.

The pathological events that underpin DGM atrophy are not known, but this is generally interpreted as the result of neurodegeneration. Previous studies have shown that DGM atrophy is more severe in patients with progressive MS, longer disease duration and worse cognitive performance (Zivadinov *et al.*, 2012). My *post-hoc* analyses showed that the thalamus, which is the DGM's largest component, was a better predictor of future disability than other regions, and the rate of atrophy in the putamen was the highest across DGM nuclei. Previous studies, including those using advanced MRI, have found that thalamic damage at study entry was associated with higher disability (Schoonheim *et al.*, 2015). DGM structures are extensively connected with cortical GM regions, and therefore DGM atrophy could be due to retrograde and anterograde neurodegeneration via tracts that connect GM areas. For

example, the extent of cellular density loss in the thalamus, is associated with neurodegeneration in the remote (but connected) cortical regions, over and beyond the extent of atrophy explained by demyelination in connecting tracts (Kolasinski *et al.*, 2012). There is also evidence of other neurodegenerative mechanisms in the DGM nuclei. For example, their higher load of iron than other regions can accumulate oxidised lipids which are associated with neurodegeneration (Hametner *et al.*, 2013). In the healthy controls, the rate of DGM atrophy was faster than that in other regions, suggesting that it may be a hot spot for both age- and disease-related atrophy in the human brain, although a methodological issue, related to its more uniform structure than other brain regions, cannot be excluded. In AUC analysis I found that at the individual level, DGM volume lacks prognostic value, which is due to the high variability typical of volumetric MRI studies (Barkhof, 2016). Nevertheless, the DGM volume holds strong promise as a marker of disease progression (at the group level) with the potential to respond to neuroprotective treatments that target neurodegeneration in MS.

Interestingly, the temporal lobe showed a significant acceleration in SPMS when compared to both RRMS and CIS. Similarly, the parietal lobe GM showed a significant acceleration of atrophy in SPMS in comparison with CIS. The *post-hoc* analysis showed that the temporal pole and insula were the most affected structures in the temporal GM. Pathological studies have demonstrated an increase in the rate of neurodegeneration, especially in the temporal regions, during progressive stages of MS in comparison with RRMS and CIS (Howell *et al.*, 2011; Haider *et al.*, 2016a). Overall, a global pathological process in MS (Lindberg *et al.*, 2004), may become more pronounced in certain regions, such as the temporal GM, because of other mechanisms, such as static exposure to CSF (the insula in the temporal lobe) or hypoxia in watershed areas (some DGM nuclei such as the pallidum). For example, meningeal inflammation and cortical demyelination, which may play a role in cortical atrophy, preferentially affect deep sulci, such as the insula, where there is more exposure to static inflammatory cytokines (Mahad *et al.*, 2015). My findings also suggest that regions with more connections may be vulnerable to atrophy. For example, among the parietal cortical regions, the precuneus, a core part of an important functional brain network (default mode network),

showed the fastest atrophy rates in SPMS (Zhang and Raichle, 2010). Thus, acceleration of atrophy during SPMS may be explained by cortical network collapse with advancing of degeneration from initial injury sites (focal lesions in the white matter or initial DGM degeneration) to interconnected neocortical systems (Seeley *et al.*, 2009). I found that MS phenotypes shared a common spatiotemporal pattern of volume loss (no significant 3-way interaction of time \times region \times phenotype). This shows, in line with previous studies, that the difference in pathology of progressive MS is only quantitative rather than qualitative in comparison with RRMS (Kutzelnigg *et al.*, 2005).

Cortical GM atrophy was seen at study entry across clinical phenotypes, even in CIS, when compared with HCs, and was the greatest in progressive MS, in agreement with earlier studies (Ceccarelli *et al.*, 2008; Audoin *et al.*, 2010). My findings of faster whole brain atrophy in SPMS, PPMS, RRMS than CIS, who in turn, showed higher cortical atrophy than HCs, are similar to previous studies on longitudinal whole brain atrophy (Kalkers *et al.*, 2002; Lukas *et al.*, 2010), regional atrophy (Riccitelli *et al.*, 2011; Lansley *et al.*, 2013; Mallik *et al.*, 2015), and pathology of MS phenotypes (Howell *et al.*, 2011). My study confirms previous findings that relationships between whole brain atrophy and clinical changes are weak or absent (De Stefano *et al.*, 2010), and shows DGM atrophy as a stronger marker of clinical disability. Although the GM volumes of cortical lobes could not predict future EDSS progression, the more detailed *post-hoc* analyses showed that regional volumes, such those of the hippocampus and the angular gyrus, were associated with future EDSS progression. These regions are highly connected to other regions, and especially the angular gyrus (like the precuneus) acts as a hub in the default mode network, which could make it vulnerable to atrophy, as explained above (Zhang and Raichle, 2010).

This study was not designed to assess the effect of treatment on atrophy rates, but does study atrophy while adjusting for possible confounding effects. The rates of atrophy in all clinical phenotypes were similar in people who were receiving disease-modifying treatments to those who were not. Even though I could not ascertain the duration of treatments due to retrospective nature of this study, the majority (90%) of patients on disease-modifying

treatments, were receiving first-line injectable drugs (interferon or glatiramer acetate) before study entry. The effects of these drugs on brain atrophy are modest at best (Filippi *et al.*, 2004; Kappos *et al.*, 2016). Therefore, drug effects are unlikely to be confounders of my analysis.

One strength of my study is that I included a large number of patients, who underwent the same protocol on the same MRI scanner over time at single sites. However, different MRI protocols could have an effect on atrophy measures and is a limitation of my study (Bendfeldt, Hofstetter, *et al.*, 2012; Biberacher *et al.*, 2016). I, therefore, used a hierarchical statistical design based on scanner. My study was powerful enough because the effects of clinical phenotype on the regional rates of atrophy were higher than the effects of between-centre variation.

I chose GIF software to segment and parcellate the brain (Cardoso *et al.*, 2015) because it allowed inclusion of 2D MRI data (which I had for one centre), and did not require any manual editing, unlike Freesurfer, which would have been unfeasible for such large number of scans. The reliability analysis showed excellent agreement between GIF-derived DGM volume and that obtained using FSL-FIRST, and between GIF-derived cortical volumes and those obtained using SPM12, respectively. Therefore, I chose to present the results obtained with GIF because it allowed us to rely on only one method to segment DGM and cortical GM, and estimate TIV. I used TIV to adjust for variations in head size, rather than the skull-size, so that a more reliable estimate of head size is obtained, irrespectively of the field-of-view, the choice of the inferior cut-off of the brain for the analysis, and demographic factors (e.g., age, or weight) (Malone *et al.*, 2015). With regard to the statistical methods, I used mixed-effects models to calculate atrophy rates (Pinheiro and Bates, 2009), which naturally accommodated multiple (3 or more) time-points with varying intervals between follow-ups, and patients who convert from one phenotype to another (e.g., CIS to RRMS). These two issues are cumbersome to address with methods that rely on pairwise comparisons (e.g., SIENA, BSI) and suffer from higher variance in brain atrophy estimates as the interval between two scans increases (Smith *et al.*, 2002, 2007). Mixed-effects modelling, instead, estimates a variance

component to eliminate implausible inconsistencies (Frost *et al.*, 2004; Cash *et al.*, 2015). Based on my experience and the results of this study, I recommend the acquisition of high-resolution 3D-T1-weighted images (isotropic 1mm³). Several methods can calculate DGM volumes, such as FSL-FIRST, and GIF. I recommend the use of the GIF software when it is desirable to use the same method to segment both the cortex and DGM.

There were also limitations in this study. The majority of centres did not provide MRI scans of HCs, however, I included a large number of HCs including those from an external initiative (PPMI). My findings of volume changes in HCs were consistent with the literature. Meta-analyses have shown, in individuals less than 70 years of age, rate of whole brain loss ranges from 0 to -0.5 (my study = -0.04), GM loss ranges from 0 to -0.5% per year (cortical GM in my study = -0.34%) (Hedman *et al.*, 2012), and the subcortical structures may show loss of up to -1.12% (DGM in my study = -0.94) (Fraser *et al.*, 2015). Cognitive functions were not tested, and it is unknown whether cortical patterns of GM atrophy over time were associated with cognitive impairment. Clinical trials in MS (and in progressive MS in particular) include confirmed disability progression, based on the EDSS, as primary outcome measure. Although for EDSS the model-estimated coefficients and their *p*-values and confidence intervals are valid for comparison between brain regions, the absolute value of these coefficients must be interpreted with caution, because the EDSS does not have a uniform linear interpretation. Since this was a retrospective study, the duration of treatments before entry to the study could not be ascertained for all participants. Disease-modifying drugs may have lasting effects, for example they may slow the accrual of disability after a decade (Kappos *et al.*, 2016; University of California, San Francisco MS-EPIC Team: *et al.*, 2016). Moreover, MRI sequences sensitive to cortical lesions were not available, and the effects of cortical lesions on atrophy measures remain unknown.

In conclusion, the DGM atrophy showed the most rapid development over time— extending previous cross-sectional studies that showed a relationship between DGM atrophy and disability— was most closely associated with disability accumulation and predicted the time to

EDSS worsening. In phase 2 trials of neuroprotective medications in MS, DGM atrophy measures may therefore have greater potential to show treatment effects than other regional GM or whole brain measures. There was a disconnect between DGM atrophy and cortical atrophy rates. The temporal and parietal cortices showed a faster rate of atrophy in SPMS than RRMS and/or CIS, whilst DGM showed a faster rate of atrophy in SPMS than CIS only, suggesting that neurodegeneration in GM regions may proceed at a different rate which should be taken into account in the design of clinical trials. Although the results of this chapter underline the differential vulnerability of specific brain structures, the sequence at which brain regions atrophy remains unclear. In the next chapter I will apply a probabilistic data-driven model to identify the underlying sequence of regional progression of atrophy in MS.

5. Progression of regional grey matter atrophy in multiple sclerosis

5.1. Summary

Grey matter atrophy is present from the earliest stages of multiple sclerosis, but its temporal ordering is poorly understood. I aimed to determine the sequence in which grey matter regions become atrophic in multiple sclerosis and its association with disability accumulation.

In this longitudinal study, I included 1,417 subjects: 253 with clinically-isolated syndrome, 708 relapsing-remitting multiple sclerosis, 128 secondary-progressive multiple sclerosis, 125 primary-progressive multiple sclerosis, and 203 healthy controls from 7 European centres. Subjects underwent repeated magnetic resonance imaging (total number of scans 3,604); the mean follow-up for patients was 2.41 years (standard deviation=1.97). Disability was scored using the Expanded Disability Status Scale. I calculated the volume of brain grey matter regions and brainstem using an unbiased within-subject template. I used an established data-driven event-based model to determine the sequence of occurrence of atrophy and its uncertainty. I assigned each subject to a specific event-based model stage, based on the number of their atrophic regions. I used linear mixed-effects models to explore associations between the rate of increase in event-based model stages, and T2-weighted lesion load, disease-modifying treatments, comorbidity, disease duration and disability accumulation.

The first regions to become atrophic in clinically-isolated syndrome and relapse-onset multiple sclerosis patients were the posterior cingulate cortex and precuneus, followed by the middle cingulate cortex, brainstem and thalamus. A similar sequence of atrophy was detected in primary-progressive multiple sclerosis with the involvement of the thalamus, cuneus, precuneus, and pallidum, followed by the brainstem and posterior cingulate cortex. The cerebellum, caudate and putamen showed early atrophy in relapse-onset multiple sclerosis and late atrophy in primary-progressive multiple sclerosis. Patients with secondary-progressive multiple sclerosis showed the highest event-based model stage (the highest number of atrophic regions, $p<0.001$) at the study entry. All multiple sclerosis phenotypes, but clinically-isolated syndrome, showed a faster rate of increase in the event-based model stage than healthy controls. T2-weighted lesion load and disease duration in all patients were

associated with increased event-based model stage, but no effects of disease-modifying treatments and comorbidity on event-based model stage were observed. The annualised rate of event-based model stage was associated with the disability accumulation in relapsing-remitting multiple sclerosis, independent of disease duration ($p<0.0001$).

The data-driven staging of atrophy progression in a large multiple sclerosis sample demonstrates that grey matter atrophy spreads to involve more regions over time. The sequence in which regions become atrophic is reasonably consistent across multiple sclerosis phenotypes. The spread of atrophy was associated with disease duration and with disability accumulation over time in relapsing-remitting multiple sclerosis.

5.2. Introduction

As I showed in the previous chapter, brain atrophy measured by magnetic resonance imaging, develops at a faster rate in people with multiple sclerosis than healthy controls. Grey matter atrophy is not uniform across the brain in multiple sclerosis, and some regions are more susceptible to atrophy than others (Steenwijk *et al.*, 2016; Preziosa *et al.*, 2017). The limbic system, temporal cortex and deep grey matter show rapid atrophy in patients with relapse-onset multiple sclerosis (Audoin *et al.*, 2010), while the cingulate cortex shows early atrophy in primary progressive multiple sclerosis (Eshaghi *et al.*, 2014). In the previous chapter, using the same large cohort of multiple sclerosis patients, I have found that the deep grey matter showed the fastest annual rate of tissue loss in relapsing-remitting multiple sclerosis and progressive multiple sclerosis, and that in the cortex the rate of atrophy accelerated in the temporal regions in secondary progressive multiple sclerosis. However, it is unknown whether there is a consistent and identifiable order in which atrophy progresses affecting different areas over time. A key question is whether there is an association between the sequential development of atrophy and disability accumulation.

One approach to investigate the sequence of atrophy progression is to employ a probabilistic data-driven method, such as an event-based model, which, as the name implies, identifies the sequence of events at which a biomarker becomes abnormal, using cross-sectional or longitudinal observations (Fonteijn *et al.*, 2012; Young *et al.*, 2014). The event-based model is an established method. It has given new insights into the progression of Alzheimer's disease in which the hippocampal atrophy is seen before the whole brain atrophy. Similarly, in Huntington's disease, the event-based model has detected the earlier atrophy in the basal ganglia than other regions.

In this chapter, I have introduced a novel validation technique for the event-based model and then used it to investigate the progression of brain atrophy as a sequence of "events" at which grey matter regions become atrophic in all phenotypes of multiple sclerosis. To define when the volume of a region ceases to be normal and becomes atrophic, the event-based model

does not rely on *a priori* thresholds but calculates the probability of atrophy based on data-derived model distributions of normal and atrophic regional volumes. Moreover, the event-based model constructs a subject staging system: it assigns each subject to a stage that reflects how far through the sequence of regions that subject shows lower than normal volumes – the higher the stage, the higher the number of atrophic areas.

In this chapter, I built on the evidence from the previous chapter and literature that neurodegeneration in multiple sclerosis does not affect all the grey matter regions equally and those brain regions become atrophic in a non-random manner. I hypothesised that: (i) there is a sequence in which grey matter regions become atrophic; (ii) this sequence differs between relapse- and progressive-onset multiple sclerosis phenotypes; and (iii) the event-based model stage increases with disease duration and disability worsening.

5.3. Methods

5.3.1. Participants

Participants are the same as those reported in Chapter 4.

5.3.2. MRI data and analysis

I collected 3D T1-weighted scans, in addition to T2-weighted/FLAIR imaging, from all centres except one. Details of the 13 different magnetic resonance imaging protocols were discussed in the previous chapter. Moreover, I used the volumes of different brain regions from the image analysis as explained in the previous chapter (**Chapter 4**).

5.3.3. The event-based model

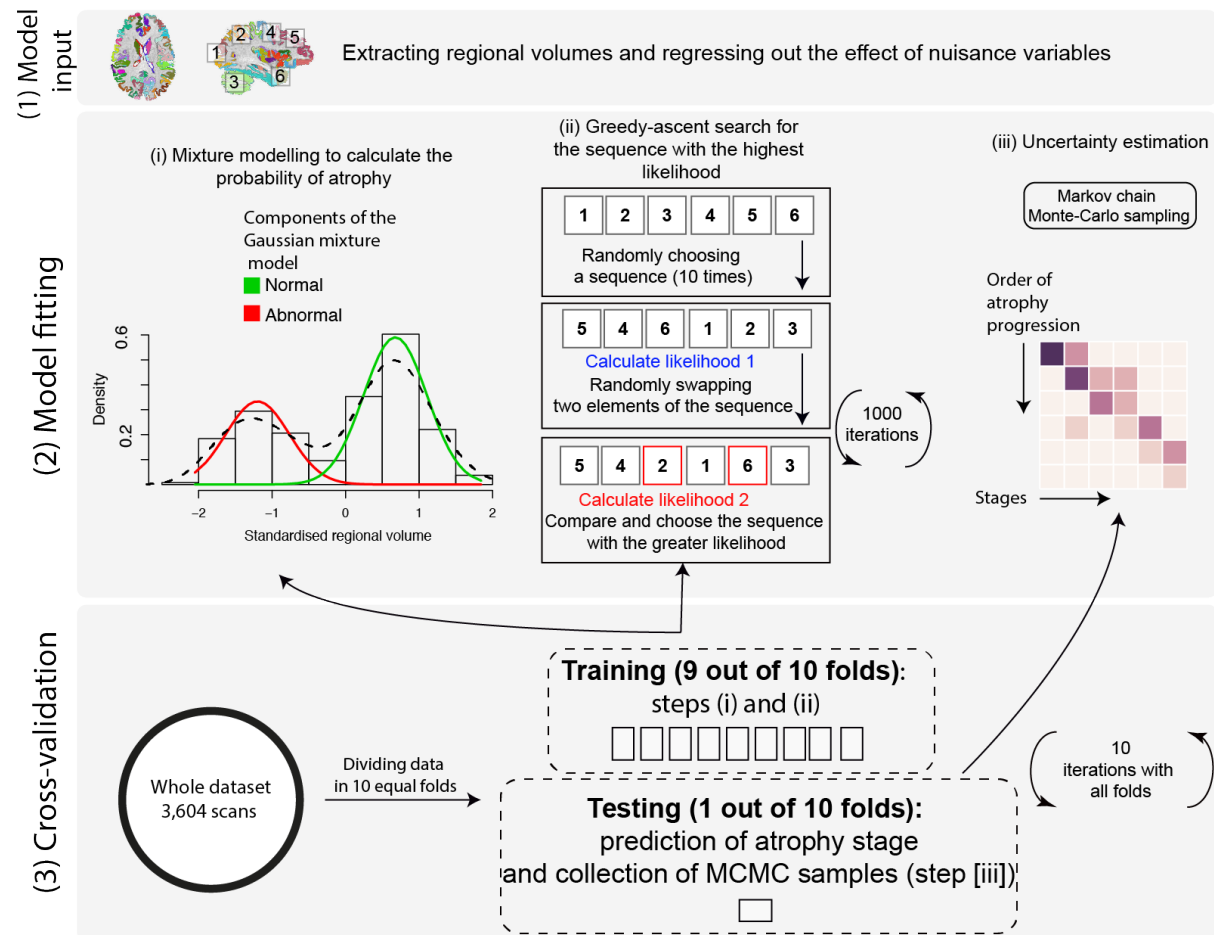
I used the event-based model, as described previously in , to estimate the most likely sequence in which selected regions become atrophic over time (see below for details on region selection). I also repeated the same analysis using all brain regions to test the dependence of my findings on the region selection.

The event-based model assumes that a population of patients represents the whole trajectory of disease progression and reconciles cross-sectional or short-term longitudinal data into a picture of the entire disease course. I, therefore, created separate event-based models for (1) relapse-onset patients (the clinically isolated syndrome, relapsing-remitting, and secondary-progressive multiple sclerosis); (2) progressive-onset (or primary-progressive) patients; and (3) all clinical phenotypes together (to develop a unique staging system for the whole cohort). I used the sequence estimated by the latter event-based model to stage patients by assigning them the most probable stage along the sequence.

The main steps of the event-based model include (**Figure 5-1**): (1) model input, which consists of the adjustment of regional volumes for effects of nuisance variables and selection of regions; (2) model fitting; and (3) a cross-validation. For the last step, I used a novel cross-validation method, used here within the event-based model for the first time, while steps one and two have not changed since the original event-based model implementations. Model input used all multiple sclerosis patients. Model fitting and cross-validation were repeated three

times using (1) relapse-onset and the clinically isolated patients together, (2) primary-progressive multiple sclerosis, and (3) the whole cohort of patients.

Figure 5-1. Estimating the most likely sequence of atrophy progression.



The event-based model steps to estimate the most likely sequence of atrophy progression. The 3 steps are: (1) adjusting for nuisance variables, and region selection (2) calculating the best-fit probability distributions for normal and atrophic brain regions; searching for the most likely sequence; and (3) quantifying the uncertainty with cross validation. (i) shows the distribution of the volume in an example region in healthy controls and patients and the corresponding mixture model. (ii) shows the steps for greedy ascent search. (iii) a matrix showing a sequence of atrophy progression on the y-axis, and the position in the sequence of each region ranging from 1 to the total number of regions on the x-axis. The intensity of each matrix entry corresponds to the proportion of Markov Chain Monte Carlo samples of the

posterior distribution where a certain region of y-axis appears at the respective stage of x-axis.

5.3.4. Model input

I adjusted the regional volumes for the total intracranial volume, age at study entry, gender, scanner magnetic field and magnetic resonance imaging protocol. Since some centres provided data from more than one imaging protocol (see previous chapter), I adjusted for imaging protocol and magnetic field (instead of “centre”). I constructed a regression model for each region separately, entering the volume as the dependent variable and the remaining variables as predictors. I extracted the amount of each regional volume that remained unexplained in the regression (residual of the fit). Subsequently, I selected the regions whose adjusted volumes at the study entry showed a significant difference between all multiple sclerosis patients and healthy controls, with a Bonferroni corrected $p < 0.01$ (non-corrected $p < 0.0001$). I used these regions in the subsequent analyses. I then repeated the analysis using all the segmented regions of the Desikan-Killiany-Tourville atlas for the following reasons: 1) to test whether the sequence in which brain regions become atrophic was not influenced by restricting the analysis only to the regions that showed a lower volume in patients than controls; 2) to detect potential subtle early changes that might not have survived multiple-comparison correction.

5.3.5. Model fitting

The event-based model considers an “event” to have occurred when a biomarker, here regional volume, has abnormal value (“atrophy”) in comparison with the expected values measured in healthy controls. The model then estimates the sequence $S = S(1), S(2), \dots, S(l)$ in which regions become atrophic, where $S(1)$ is the first region, and $S(l)$ is the last to become atrophic. The model assumes that all patients go through the same sequence as they progress. The estimation procedure first fits a mixture of two Gaussians to regional volumes,

with one of the components fixed to be identical to the healthy distribution; the other component provides the model for the “abnormal” distribution. This provides probabilistic models for normal and abnormal volumes from which we can calculate the likelihood of atrophy $P(x_{ij}|E_i)$ for the region i of the scan j , i.e. the probability density function (PDF) estimated at x_{ij} from the abnormal component of the mixture model. The likelihood that region i has no atrophy or $P(x_{ij}|\neg E_i)$, is the PDF of the normal component of the mixture-model estimated at x_{ij} (see **Figure 5-1**, section 2[i]).

To search for the most likely sequence, I used a greedy ascent search (Young *et al.*, 2014) which started at ten different random sequences and iterated by randomly flipping sequences for 1000 times. The final sequence was selected when ten different initial sequences converged to a similar likelihood after 1000 iterations. Within each iteration new (flipped) sequences (Figure 1, section 2[ii]) were accepted only if they increased the likelihood, which is defined as

$$P(X|S) = \prod_{j=1}^J \left[\sum_{k=0}^I \left(P(k) \prod_{i=1}^k P(x_{ij}|E_i) \prod_{i=k+1}^I P(x_{ij}|\neg E_i) \right) \right] \quad (1)$$

where X is the data matrix, S is the sequence of atrophy events, J is the number of scans, I is the number of regions, and $P(k)$ is the prior probability of being at stage k , which means E_1, \dots, E_k have occurred, and E_{k+1}, \dots, E_I have not occurred. I used a uniform distribution for prior probabilities, which assumes equal prior-probability for all possible stages; all sequences are equally likely *a-priori*. The software and codes for the event-based model are freely available at <https://github.com/ucl-mig/ebm>.

5.3.6. Cross-validation of atrophy sequence

After estimating the most likely sequence, the uncertainty in the position of each region in the sequence was estimated using cross-validation and Markov Chain Monte-Carlo. I divided the dataset (including baseline and follow-up visits) into ten equally-sized folds (cross-validation folds) and repeated the sequence estimation ten times. During each iteration, I used nine-folds to fit the mixture-models (as explained above) and estimated the most-likely sequence. I kept one fold out as the test fold to assign the event-based model stages (explained below).

Within each iteration, I used Markov Chain Monte-Carlo to sample from the posterior distribution on the sequence given the nine-fold training data, as in (Fonteiin *et al.*, 2012; Young *et al.*, 2014). I then aggregated Markov Chain Monte-Carlo samples from the ten iterations of cross-validation (10,000 samples from each fold) to calculate uncertainty across cross-validation folds. Finally, I used these 100,000 sampled sequences to plot the positional variance diagram [as in (Fonteiin *et al.*, 2012; Young *et al.*, 2014)], which shows on the y-axis the sequence with the highest likelihood, and the x-axis enumerates the number of sequence positions (or the event-based model stages). The intensities of the matrix entries correspond to the proportion of Markov Chain Monte-Carlo samples in which the corresponding region (y-axis) appears at the respective stage (x-axis). Therefore, if there were no uncertainty, i.e. all Markov Chain Monte-Carlo samples in all folds find the same sequence, the matrix would be black on the diagonal and white everywhere else; non-white off-diagonal and non-black diagonal elements indicate uncertainty in the position of the corresponding region in the sequence.

5.3.7. Staging individual subjects and associations with white matter lesion load, disease duration and disability

I used the most likely sequence of atrophy progression from the whole patient cohort-based event-based model to obtain the event-based model stage for each scan j , which is the stage k that maximises $\prod_{i=1}^k P(x_{ij}|E_i) \prod_{i=k+1}^l P(x_{ij}|\neg E_i)$. This assigned each subject an event-based model stage between 1 and the number of regions, l , at each visit (see **Figure 5-1**).

I used a nested linear mixed-effects model to investigate the association between the event-based model stage (dependent variable) and T2-weighted lesion load (independent variable), in which time was nested in subject as the random-effect (to adjust for repeated measures). Similarly, I used a nested mixed-effects regression model to explore the association between the event-based model stage (dependent variable) and disease duration (independent variable), in which disease duration was nested in subject as the random-effect.

For those clinical phenotypes that showed a significant change in the event-based model stage over time (relapsing-remitting, secondary-progressive and primary-progressive multiple sclerosis), I investigated whether longitudinal Expanded Disability Status Scale changes could be predicted by event-based model changes independent of disease duration. I divided the changes in the Expanded Disability Status Scale and event-based model by the number of years from the study entry and performed a linear regression analysis where the annualised Expanded Disability Status Scale change was the outcome variable. Annualised event-based model stage change and disease duration at the study entry were the predictor variables. Since both the event-based model stage and Expanded Disability Status Scale are ordinal variables, I used ordinal regression analyses to confirm the results of the linear regressions but presented the results of linear models (as they did not materially differ).

5.3.8. Confounding effects of disease-modifying treatments and comorbidities

To test whether disease-modifying treatments could affect the event-based model stages at baseline and over time, I used similar mixed-effects models (as above) in which the event-based model stage was the outcome variable; time, disease-modifying treatment (as a categorical variable), and their interaction were the fixed-effect variables. Random-effects were the same as explained above. I performed additional analyses to assess the effects of comorbidities on the event-based model stages, which are reported in the Supplementary Material.

Effects of comorbidities on atrophy measures

I performed a subgroup analysis on the subset of patients with available comorbidity information. The comorbidity was defined as a categorical variable that showed the documented presence (or absence) of diabetes, hypertension, hyperlipidaemia, ischemic heart disease, stroke, other autoimmune diseases, psychiatric diagnosis (major depression disorder, bipolar disorder, alcohol abuse, and other psychiatric disorder), chronic lung disease, renal disorders, migraine, and smoking before the date of first scan (Geraldes *et al.*, 2017; Marrie, 2017). I aimed to compare the effect of comorbidities on the rate of change in the

event-based model stage. I used a linear mixed effects model in which the event-based model stage was the outcome variable. Time, comorbidity, and the interaction of comorbidity with time were the fixed-effect variables. Time was nested in the centre as the random-effect.

5.4. Results

5.4.1. Subject characteristics

Imaging data from 1,424 subjects (the image analysis is explained in **Chapter 4**) were analysed; three subjects' scans were excluded because of motion artefacts and four because of poor registration due to missing imaging header information. Therefore, data from 1,417 subjects were included in the final modelling: 1,214 patients (253 clinically isolated syndrome, 708 relapsing-remitting, 128 secondary-progressive, and 125 primary-progressive multiple sclerosis), and 203 healthy controls. The average (\pm standard deviation) length of follow-up for patients was 2.43 years (\pm 1.97) and for healthy controls was 1.83 years (\pm 1.77). In total, I analysed 3,604 T1-weighted scans (mean number of scans per patient was 2.54 [SD=1.04]).

Table 5-1. Baseline characteristics of participants.

Group	Healthy control	Clinically isolated syndrome	Relapse onset multiple sclerosis*	Primary-Progressive multiple sclerosis
Total number (number of females)	203 (112)	253 (171)	836 (548)	125 (55)
Age (\pm SD ¹)	38.7 \pm 10.5	33 \pm 8	39.7 \pm 9.8	48.5 \pm 10.1
Disease duration (\pm SD ¹)	—	0.4 \pm 1.4	8.06 \pm 8.03	6.8 \pm 5.9
Median Expanded Disability Status Scale (range)	—	1 (0-4.5)	2 (0-9)	5 (2-8)
Percent (number) of patients receiving disease-modifying treatments	—	20% (52)	47% (397)	6% (8)
Baseline median white matter T2 lesion load (ml) (1 st -3 rd quartile)		2.97(1.01-5.04)	5.04 (2.05-11.79)**	9.38 (2.69-22.02)

¹ standard deviation

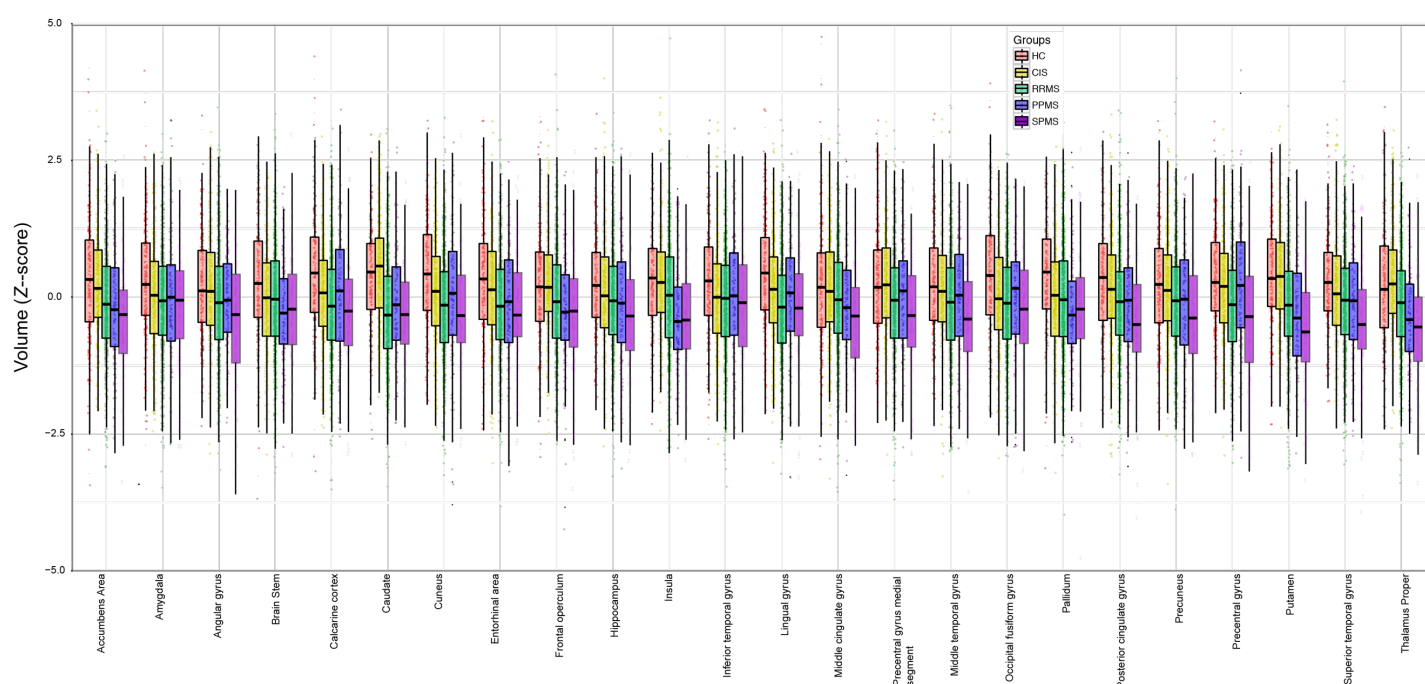
* Relapse onset group includes both the relapsing-remitting and secondary progressive patients.

** Baseline median T2-weighted lesion loads were the following: for relapsing-remitting=5.05 (2.05-11.79) and secondary-progressive multiple sclerosis =11.04 (3.18-23.14).

5.4.2. Sequence of atrophy progression

At baseline, 24 regions showed a smaller volume in multiple sclerosis than healthy controls (Bonferroni corrected $p < 0.01$). They included the deep grey matter regions and the posterior cortices (including the precuneus and the posterior cingulate cortex), several regions in the temporal lobe, the precentral cortex, and the brainstem (see **Figure 5-2** for the full list).

Figure 5-2. Comparisons of regional volumes between groups.

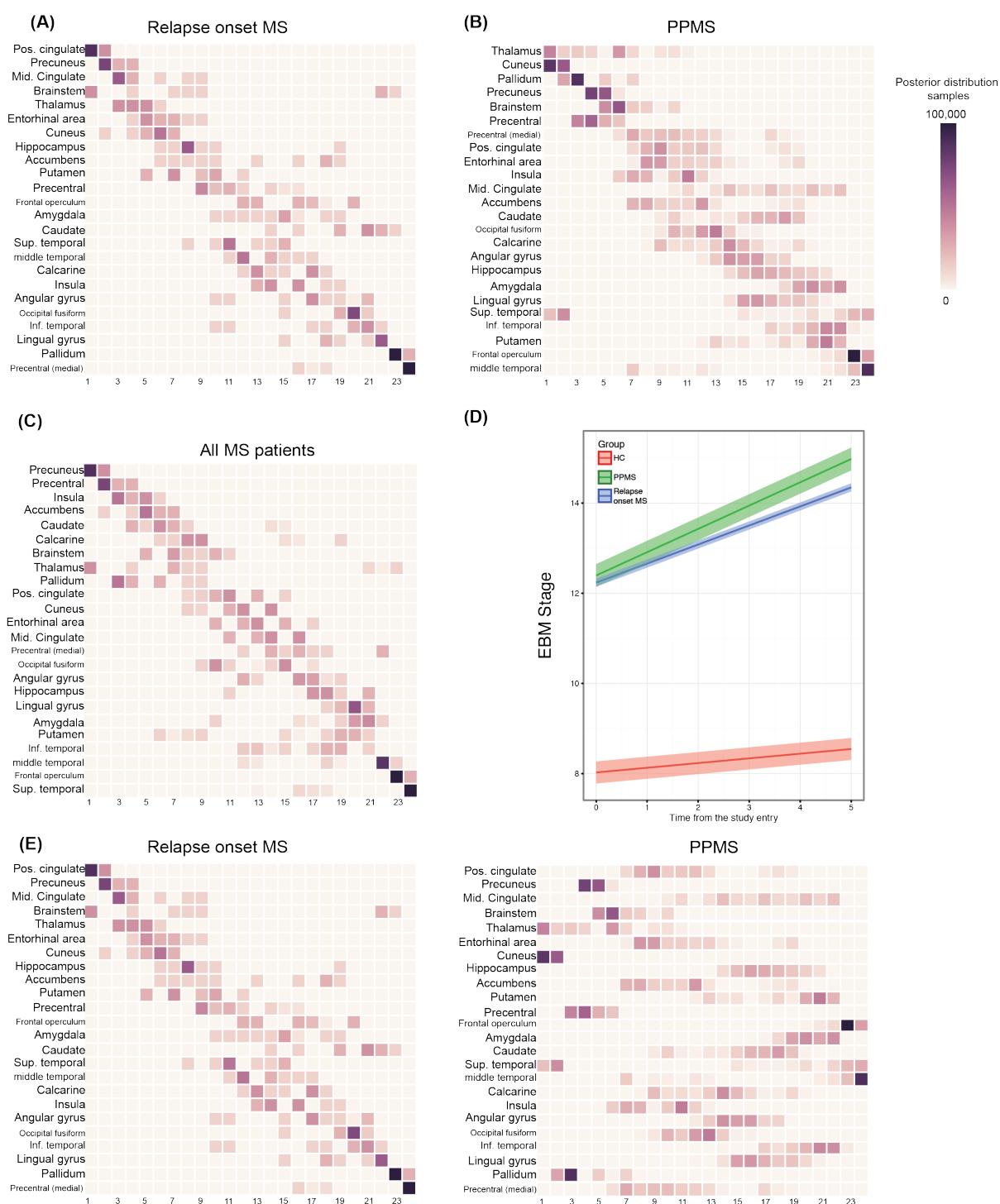


Box plots at y-axis show z-scores of the corresponding region shown at x-axis. Lower and upper hinges of each boxplot correspond to 25th and 75th percentiles of data. I selected 24 regions that showed significant difference ($p < 0.01$ corrected) between all patients with MS and healthy controls at the baseline visit.

When I estimated the sequence in which these 24 regions become atrophic in patients with relapse-onset multiple sclerosis (i.e., relapsing-remitting and secondary-progressive) and the clinically isolated syndrome, the first regions were the posterior cingulate cortex and precuneus, followed by the middle cingulate cortex, brainstem, and thalamus (**Figure 5-3A & D**); the last regions to become atrophic were the pallidum and medial precentral gyrus.

In patients with primary-progressive multiple sclerosis, among the 24 selected regions, the first ones to show atrophy were the thalamus, cuneus and precuneus, and pallidum, followed by the brainstem, precentral gyrus, and posterior cingulate cortex (**Figure 5-3B & D**); the last regions to become atrophic were the frontal operculum and middle temporal gyrus.

Figure 5-3. Sequences of atrophy progression and patient staging.



The positional variance diagrams for (A) relapse-onset multiple sclerosis, (B) primary progressive multiple sclerosis and (C) merged cohort of patients, show the most likely sequences of atrophy and their associated uncertainty. In (A), (B), and (C) the y-axis shows the most likely sequence of atrophy progression, and the x-axis shows the sequence position

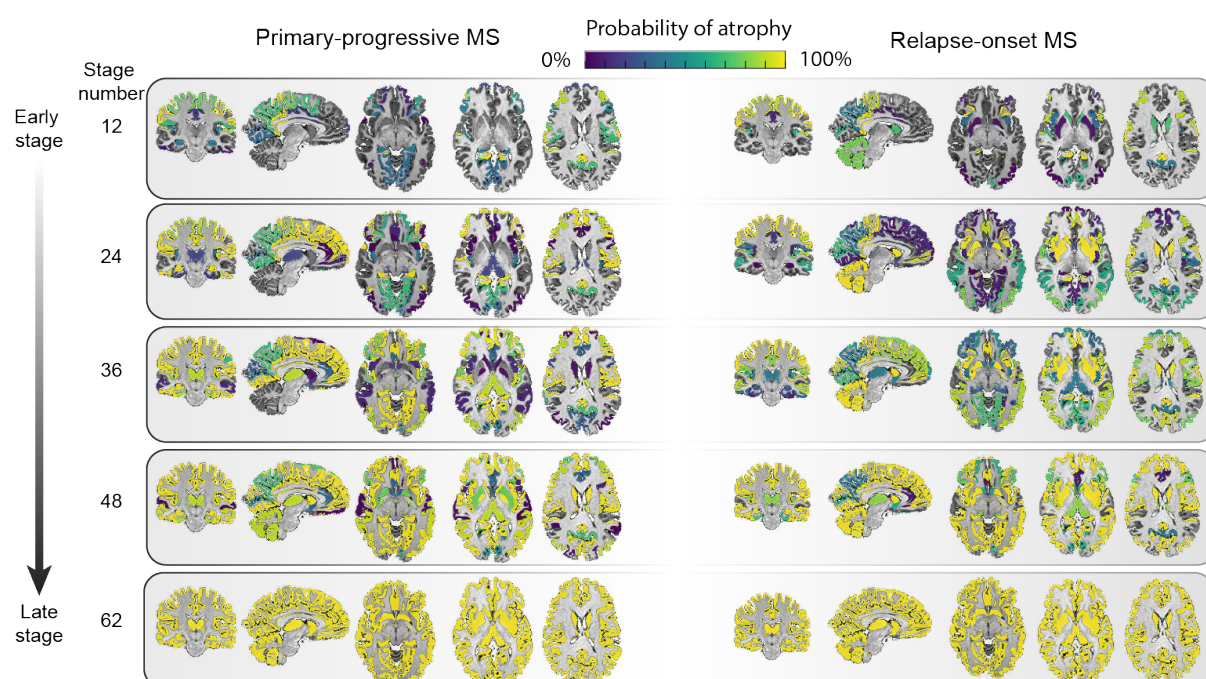
ranging from one to the total number of regions. The intensity of each rectangle corresponds to the proportion of Markov Chain Monte Carlo samples of the posterior distribution where a certain region of y-axis appears at the respective stage of x-axis. (D) shows the evolution of the event-based model stage (or atrophy progression staging) over time in clinically isolated syndrome and relapse-onset multiple sclerosis together and primary progressive multiple sclerosis. Each line in (D) is the prediction of mixed-effects model whose ribbon shows standard error of the prediction. (E) shows the same positional variance diagram as (A) and (B) with the only difference that the rows of the relapse-onset positional variance diagram have been re-ordered to have the same order of primary progressive multiple sclerosis to facilitate the comparison.

When the event-based model was used to estimate the sequence of atrophy progression of the selected 24 regions in all patients together, additional regions were detected as showing early atrophy, such as the insula, accumbens and caudate (**Figure 5-3C**). The likelihood of the ten randomly chosen sequences (log-likelihood range: -149000 to -117000) converged to a similar range (log-likelihood range: -1000000 to -99000) after 1000 iteration (**Supplementary Figure 5-1**). For other event-based models, the likelihoods converged to a similar range (results are not shown).

When all the remaining regions were included additional regions were identified. In primary-progressive multiple sclerosis, they were the transverse temporal gyrus, cerebral white matter, post-central gyrus and middle frontal gyrus (see **Figure 5-4, Supplementary Figures 5-2 and 5-3**). In the relapse-onset group, these regions were the superior frontal gyrus, inferior frontal gyrus, and middle frontal gyrus.

When I qualitatively compared the clinically isolated syndrome and relapse-onset multiple sclerosis patients with primary-progressive multiple sclerosis, across all regions, the cerebellum, caudate and putamen showed a differential pattern of atrophy, with early atrophy in patients with relapse-onset disease and late atrophy in primary progressive multiple sclerosis (see **Figure 5-4**).

Figure 5-4. Regional atrophy and its sequence of progression in all grey matter regions plus brainstem in relapse-onset disease and primary progressive multiple sclerosis.



The probability of atrophy in each region was calculated from the positional variance diagrams and colour coded, so that brighter colour corresponded to higher probability of seeing atrophy in the corresponding event-based model stage.

5.4.3. Event-based model staging of individual subjects

Patients with clinically isolated syndrome and relapse-onset multiple sclerosis and primary-progressive multiple sclerosis had significantly higher event-based model stages at baseline than healthy controls (average intercept [\pm standard error] of the event-based model stage for HCs=8.02 [\pm 0.59], relapse-onset=12.39 [\pm 0.66], primary-progressive multiple sclerosis=12.22 [\pm 0.35], $p<0.05$); when looking at each clinical phenotype, patients with secondary-progressive multiple sclerosis had the highest event-based model stage at the study entry (14.73 [\pm 0.93], all p -values <0.001), followed by relapsing-remitting (12.60 [\pm 0.67]), primary-progressive multiple sclerosis (12.22 [\pm 0.35]), clinically isolated syndrome (8.12 [\pm 0.76]), and healthy controls (8.02 [\pm 0.59]). The annual rate of change (or slope) in the event-based model stage over time was significant (null-hypothesis=zero slope) for secondary-progressive

multiple sclerosis (average slope [\pm standard error]=1.02 [\pm 0.41]), primary-progressive (0.52 [\pm 0.34]), and relapsing-remitting multiple sclerosis (0.37 [\pm 0.26]), but not for clinically isolated syndrome (0.19 [\pm 0.33]) and healthy controls (0.10 [\pm 0.24]). The rate of change, although nominally higher in secondary progressive multiple sclerosis, was not significantly different between clinical phenotypes.

5.4.4. Association of the event-based model stages with white matter T2-weighted lesion load, disease duration and disability accumulation

At baseline, the highest T2-weighted lesion load was observed in secondary progressive multiple sclerosis, followed by primary-progressive multiple sclerosis (**Table 5-1**). There was a significant association between the event-based model stage and white matter lesion load (standardised β =0.11, p <0.001) in all patients, which means for every standard deviation (15.31 millilitre) increase in the lesion load there was 0.11 standard deviation (1.06 unit) increase in the event-based model stage. However, there was no association between the rate of change in the event-based model stage over time and the rate of increase in lesion load.

There was a significant association between the rate of increase in the event-based model stages and disease duration in all patients with multiple sclerosis (β =0.21, standard error=0.03, p <0.001) using all available time points. This means that for every increase of one event-based model stage, disease duration increased by 4.76 years.

At the baseline visit, there was no significant association between the event-based model stage and the Expanded Disability Status Scale in any clinical phenotype. Over time there was a significant increase in the Expanded Disability Status Scale in both relapse-onset multiple sclerosis/clinically isolated syndrome and primary-progressive multiple sclerosis patients (an increase of 0.07 and 0.2 per year, respectively, p <0.01). There was a significant association (independent of disease duration) between the annualised event-based model stage and the annualised Expanded Disability Status Scale changes in relapsing-remitting multiple sclerosis (beta =0.03, p <0.0001), but not in secondary-progressive and primary-progressive multiple sclerosis. This means that assuming a linear relationship between the Expanded Disability

Status Scale and the event-based model stage, for every unit increase in the annual rate of the event-based model stage there is 0.03 increase in the annual rate of Expanded Disability Status Scale worsening.

5.4.5. Disease-modifying treatments and comorbidity did not affect the event-based model stages

Information on whether or not a patient was receiving a disease-modifying treatment was available for 98% of patients (n=1,179) at baseline. Of these, 38% (n=457) were receiving a disease-modifying therapy; 47% of patients with relapse-onset MS, 20% of patients with CIS and 6% of patients with primary-progressive MS were on treatment (**Table 5-1**). Information on the type of disease-modifying treatment was available for 56% of these patients (n=255), of whom 86% (n=220) were receiving either interferons or glatiramer acetate, and the remaining 14% (n=35) patients were on other treatments, including natalizumab, fingolimod, mitoxantrone, and teriflunomide. Linear mixed-effects models showed that at baseline (estimated average \pm standard error) the event-based model stage was not significantly different ($p=0.21$) between patients who were on disease-modifying treatments (12.63 ± 0.32) compared to those who were not (11.98 ± 0.52). The same model showed that the annual rate of change (estimated range \pm standard error) in the event-based model stage was not significantly different ($p=0.45$) between patients who were on disease-modifying treatments (0.53 ± 0.17) and those who were not (0.39 ± 0.10).

5.4.6. There was no effect of comorbidities on event-based model stages after accounting for age

The data on comorbidities were available for 28% of patients (n=340) at baseline. Out of these patients, 183 patients (18 primary-progressive multiple sclerosis, 110 relapsing-remitting multiple sclerosis, 22 secondary-progressive multiple sclerosis, and 33 clinically isolated syndrome), had at least one comorbidity and the remaining 157 (23 primary progressive multiple sclerosis, 86 relapsing-remitting multiple sclerosis, 8 secondary-progressive multiple sclerosis, and 40 with the clinically isolated syndrome) had no comorbidity. At baseline,

patients without comorbidity were significantly younger than those with comorbidity (37.27 ± 9.79 vs 40.2 ± 11.56 , $p=0.01$). The number of atrophic brain regions (or the event-based model stage) did not differ between patients with and without comorbidity (11.03 (9.41) vs 9.08 (9.01), $p=0.18$). Similarly, the estimated annual rate of change (\pm standard error) in the event-based model stage did not differ between the two groups (0.575 (± 0.22) vs 0.76 (± 0.32), $p=0.55$, adjusted for age).

5.5. Discussion

In this study, I used a data-driven method to determine the most likely sequence in which brain regions become atrophic in multiple sclerosis. This sequence is consistent in key regions across multiple sclerosis phenotypes: the posterior cingulate cortex, precuneus, and thalamus were among the earliest regions to become atrophic in both relapse-onset phenotypes and primary-progressive multiple sclerosis. The event-based model staging system was applied to individual patients, and the rate of increase in the event-based model stage was associated with the disease duration in all multiple sclerosis phenotypes and with the Expanded Disability Status Scale in patients with relapsing-remitting multiple sclerosis independent of the disease duration. My findings in this chapter confirm those of the previous chapter, and provide novel insights into the mechanisms of disease worsening in multiple sclerosis

The order of atrophy progression in the event-based model for most regions was similar between primary-progressive multiple sclerosis and the clinically isolated syndrome/relapse-onset multiple sclerosis. This may support the evidence from histological studies that the pathological processes are regionally consistent between early relapsing-remitting and progressive multiple sclerosis (Mahad *et al.*, 2015). These results showed that areas with an early atrophy were the posterior cingulate cortex, precuneus, thalamus and brainstem in both groups, thereby extending the results of previous studies, which have limited their investigation to specific multiple sclerosis subtypes (Gilmore *et al.*, 2009; Audoin *et al.*, 2010; Calabrese, Reynolds, *et al.*, 2015; Steenwijk *et al.*, 2016). When all patients were included together, the insula, accumbens and caudate were predicted as becoming atrophic early on. The cingulate cortex and insula have extensive connections with other regions. Possible factors for their early atrophy, therefore, can include disconnection secondary to white matter lesions, inflammation, and more specifically meningeal inflammation. I, therefore, calculated white matter T2-weighted lesion volumes and showed that there was an association between increasing lesion load at baseline and the event-based model stages. Since assessing meningeal inflammation is very challenging in vivo I can just speculate that structures in

cortical invaginations, can be exposed to meningeal inflammation, cortical demyelination, and neurodegeneration (Gilmore *et al.*, 2009; Howell *et al.*, 2011; Haider *et al.*, 2016a). The cingulate cortex and precuneus are part of a network of active regions during rest (the default mode network) (Raichle, 2015). These regions are interconnected with other areas, have the highest energy consumption in the brain, and are affected by multiple sclerosis and other neurodegenerative disorders (Bonavita *et al.*, 2011). In multiple sclerosis, neurons with demyelinated axons consume more energy to adapt to demyelination, which creates a microenvironment similar to that of hypoxia (“virtual hypoxia”) (Trapp and Stys, 2009). Neurons that survive in a state of persistent virtual hypoxia are more vulnerable to degeneration (Zhang and Raichle, 2010), and this may explain the higher vulnerability of the cingulate and precuneus cortex to atrophy.

Other regions that showed early atrophy were the thalamus and the brainstem in both relapse-onset multiple sclerosis and primary-progressive multiple sclerosis. In the previous chapter, I found that the deep grey matter showed the fastest rate of atrophy over time, while brainstem had the highest atrophy (the lowest volume) at study entry, but its atrophy progressed at a slower rate than that occurring in other regions. This may suggest that during early stages of multiple sclerosis, the rate of atrophy in the brainstem is higher than later stages, while the rate of atrophy in the thalamus remains high throughout the disease course. The brainstem is in close contact with the spinal cord, whose atrophy is seen from early stages of multiple sclerosis independent of the cortex or deep grey matter (Ruggieri *et al.*, 2015).

Several mechanisms may underlie neurodegeneration in the deep grey matter, including mitochondrial failure, iron deposition, retrograde degeneration through white matter lesions, and meningeal inflammation (for structures closer to cerebrospinal fluid) (Calabrese, Magliozzi, *et al.*, 2015; Bodini *et al.*, 2016; Pardini *et al.*, 2016). Network overload and collapse, similar to the cingulate and precuneus cortex, could also explain preferential atrophy of the deep grey matter in multiple sclerosis (Minagar *et al.*, 2013).

There were a few regions showing a differential pattern of atrophy between relapse- and progressive-onset phenotypes. The cerebellum, caudate and putamen were predicted to have

early atrophy in relapse-onset disease and late atrophy in primary-progressive multiple sclerosis. In the cerebellum, this different behaviour can be explained by a more inflammatory phenotype of patients with relapse-onset multiple sclerosis. In patients with multiple sclerosis, more than any other brain region, demyelination is seen in the cerebellar grey matter, which is five times more than the white matter demyelination (Gilmore *et al.*, 2009). This may be a consequence of overlying meningeal inflammation in the deep folia, which accommodate a static inflammatory milieu (such as cytokines, and immunoglobulins) (Kutzelnigg *et al.*, 2007; Howell *et al.*, 2011). Therefore, in the cerebellum overlying inflammation may play a role and amplify other pathological mechanisms, such as retrograde neurodegeneration secondary to white matter lesions. Thus, the cerebellum could be susceptible to inflammatory damage from the cerebrospinal fluid. Previous studies have reported in relapse-onset multiple sclerosis, but not primary-progressive multiple sclerosis, tertiary lymphatic follicles in cortical invaginations, which may suggest a more inflammatory cerebrospinal fluid milieu than primary-progressive multiple sclerosis (Kutzelnigg *et al.*, 2007; Choi *et al.*, 2012). This could explain earlier atrophy of the cerebellar grey matter in people with relapse onset disease, while in primary-progressive multiple sclerosis, neurodegeneration in a less inflammatory cerebrospinal fluid milieu might cause a gradual progression of atrophy (Choi *et al.*, 2012; Mahad *et al.*, 2015). However, this is speculative, and it remains unclear whether meningeal inflammation has a causative effect on demyelination and neurodegeneration.

The caudate and putamen, which are histologically similar, constitute a structure that is known as the neostriatum. A previous histopathological study has shown that the greatest extent of demyelination and lesions in the deep grey matter can be seen in the caudate even in early multiple sclerosis, although the pattern was not different between multiple sclerosis phenotypes (Haider *et al.*, 2014). Moreover, the putamen receives significant inputs from the motor cortex and the caudate from the association cortices. Therefore, I could speculate that retrograde neurodegeneration secondary to a higher lesion load in relapse-onset disease (compared to primary-progressive multiple sclerosis) may perform as an additive factor on demyelination to explain the higher vulnerability of these structures.

I extended my analysis from regions that showed significant atrophy at baseline to the all segmented areas to test the dependence of my findings to region selection. Another reason was to explore early, but subtle, changes in brain regions, which might have been missed by just looking at a snapshot at the study entry to choose specific areas whose adjusted volumes showed a significant difference between all multiple sclerosis patients and healthy controls, based on stringent multiple-comparison correction. For example, a brain region may show mild volume loss earlier than another part with a greater (but later) volume loss through the course of multiple sclerosis. Whole brain event-based model analysis predicted an early involvement of the posterior cortices (posterior cingulate and precuneus) along with the brainstem. New additional regions in the whole brain event-based model were also identified as showing atrophy at an early stage, including the superior, middle, and inferior frontal gyri in relapse-onset phenotypes, and the transverse temporal gyrus, white matter, and post-central gyrus in primary-progressive multiple sclerosis. These findings suggest that the changes in these structures may happen early, but with a lower intensity than other regions that were selected initially (24 areas).

This study was not designed to investigate the effects of disease-modifying drugs and comorbidities on the atrophy stages. However, it does study the sequence of regional atrophy in the presence of these confounders. There were no significant differences at baseline or during the follow-up in the event-based model stages of patients who were receiving disease-modifying treatments and those who were not, extending the results of **Chapter 4** of the same group of patients which demonstrated, using a different statistical method, that the rates of atrophy in neuroanatomical regions were not confounded by disease-modifying treatments (see **Chapter 4**, Results). As discussed in the previous chapter, most of the patients were receiving the injectable first-line therapies (interferon beta and glatiramer acetate), whose effects on atrophy rates are weak. Although the information on comorbidities was only available for about a third of patients, I found that patients with at least one comorbidity were older at baseline than those without. This is in line with the literature showing that comorbidities are prevalent in patients with multiple sclerosis and increase with age (Marrie and Horwitz,

2010; Geraldes *et al.*, 2017). There was no significant effect of comorbidity on event-based model stage at baseline or its rate of change during the follow-up. One reason that age influenced the frequency of comorbidities, but not the number of atrophic regions, was that I had regressed out the effects of age on the regional volumes. Therefore, I conclude that disease-modifying treatments and comorbidity did not significantly influence my findings.

The event-based model has a potential for clinical use as it does not rely on time and can be applied to individual (cross-sectional) brain scans. To have a unique staging system across all clinical phenotypes, I created an event-based model from the whole patient cohort. I showed that patients with secondary-progressive multiple sclerosis had the highest event-based model stage—or the highest number of atrophic regions—at the study entry. This, in line with previous studies, suggests that secondary-progressive multiple sclerosis has more advanced neurodegeneration across multiple sclerosis phenotypes (Ceccarelli *et al.*, 2008). When I performed the event-based model staging using follow-up scans of patients and healthy controls, I found a significant increase in event-based model stages in all multiple sclerosis phenotypes, but not in the clinically isolated syndrome or healthy controls (although the baseline event-based model stage was nominally higher in the clinically isolated syndrome than healthy controls). The clinical relevance of the event-based model was confirmed by a significant association between stages and Expanded Disability Status Scale in relapsing-remitting multiple sclerosis, after adjusting for disease duration. Therefore, the sequential pattern of atrophy may explain disease worsening in relapsing-remitting multiple sclerosis. I did not find the same association between the changes in event-based model stages and Expanded Disability Status Scale in other patient groups. However, patients with secondary-progressive multiple sclerosis had the highest event-based model stages at the study entry and the highest (nominal) rate of increase in the event-based model stage.

In this study I applied the current implementation of the event based model, which has been originally developed for the analysis of cross-sectional data (e.g., (Fontejn *et al.*, 2012)), although it has also been used in longitudinal analysis (Young *et al.*, 2014). This implementation has been extensively validated (Oxtoby, Alexander, *et al.*, 2017; Oxtoby,

Garbarino, *et al.*, 2017; Oxtoby *et al.*, 2018; Wijeratne *et al.*, 2018) and is feasible for clinical application because it does not require the availability of follow-up scans. However, a limitation of the current implementation of event-based model is that it does not take into account the within and between subject variance components. In this study I applied mixed-effects models to the results of event-based model stages with subject as random effects to address this. However, the explicit modelling of variance components are not included in the event-based model. Future implementations of event-based model will integrate within- and between-subject covariance components. At the time of writing of this thesis, the mixed-effects implementation of event based model is under development.

Although this is a retrospective and multi-centre study, I have adjusted for the effects of MRI protocol and scanner magnetic field, and, as reported in the previous chapter, the effect of multiple sclerosis phenotypes on regional measures was higher than that from these variables. A possible limitation is that event-based model assumes that all brain regions eventually become abnormal (all regions show atrophy at the last stage). Therefore, an implicit assumption is that patients with relapse-onset disease (the clinically isolated syndrome, relapsing-remitting, and secondary-progressive multiple sclerosis) represent the whole continuum of progression when analysed separately; future implementations of this model could remove this assumption. I used the Expanded Disability Status Scale as the clinical outcome, but both the Expanded Disability Status Scale and event-based model provide measures that are ordinal, and may not have a uniform interpretation. Therefore, the coefficients of associations should be interpreted relatively (e.g., to compare clinical groups) rather than absolutely.

I showed that the sequence of atrophy progression in relapse-onset disease and primary-progressive multiple sclerosis are similar in many key regions, while the cerebellum, caudate and putamen show an earlier atrophy in relapse-onset multiple sclerosis than primary-progressive multiple sclerosis, perhaps due to a more inflammatory milieu. The sequence of atrophy progression can be used to score patients during multiple sclerosis automatically.

6. Brain atrophy mediates the simvastatin effect on disability: a mechanistic modelling study

6.1. Summary

Neurodegeneration is the substrate of disability worsening in multiple sclerosis. A recent clinical trial has shown that brain atrophy is delayed by high-dose simvastatin in patients with secondary progressive multiple sclerosis. However, the common analysis of trial outcome measures precludes a mechanistic understanding of drug actions. Here I investigated the *causal* chain of events that link the treatment to brain atrophy and disability. I also investigated the spatiotemporal pattern of brain atrophy and whether it was altered by simvastatin. I re-analysed the double-blind, randomised controlled trial of placebo vs high-dose simvastatin (MS-STAT trial) in which 140 patients with secondary progressive multiple sclerosis underwent imaging, clinical and cognitive assessments at baseline, after one and two years. Patients who completed the trial ($n=131$) at baseline had on average (\pm standard deviation) 51.2 (± 6.97) years of age, 21.17 (± 8.65) years of disease duration, and a median Expanded Disability Status Scale of 6 (range=4-7). Participants were assessed on the cognition (Block Design test), a patient-reported outcome measure (Multiple Sclerosis Impact Scale 29), and the Expanded Disability Status Scale. I longitudinally processed anatomical scans to calculate annual percentage changes of the whole brain, cortical, and subcortical volumes with mixed-effects models. I performed multivariate mechanistic modelling to quantify interrelations of ongoing brain atrophy, with concomitant changes in the disability, Block Design test and Multiple Sclerosis Impact Scale 29. The simvastatin group showed slower rates of disability worsening than the placebo, as measured by the Expanded Disability Status Scale (beta=0.08, standard error=0.04 vs. 0.21 ± 0.03 , $p=0.002$), Block Design test (0.92 ± 0.45 , vs. -0.13 ± 0.33 , $p=0.04$) and the physical subtest of Multiple Sclerosis Impact Scale 29 (0.43 ± 0.72 vs. 2.05 ± 0.56 , $p=0.03$). The most likely model indicated that simvastatin directly slowed atrophy rates and disability. Slowing of atrophy caused an additional effect on clinical outcome

measures; 33.4% of the total treatment effect on Block Design test, and 30.9% of the effect on clinical disability were mediated by atrophy. Simvastatin reduced the annualised rate of atrophy in several brain regions, with the highest effect reaching 50% in the transverse temporal gyrus (annual rate [95% confidence interval]; placebo = -1.58% [-1.17%,-1.98%]; simvastatin=-0.79% [-0.22%,-1.35%]; $p=0.002$). In conclusion, I showed that the slowing of clinical and cognitive worsening is *caused* by a reduction in brain atrophy rate in following a cascade of events that slows the worsening of the patient reported outcome. My findings suggest that simvastatin delays brain atrophy by a global neuroprotective effect.

6.2. Introduction

An unmet need for multiple sclerosis research is the discovery of mechanisms underlying disability worsening. The substrate of disability worsening is neurodegeneration, which is captured as brain volume loss (or atrophy) on serial magnetic resonance imaging (see previous chapters for detailed discussion on this). The treatment effect on brain atrophy is correlated with the effect on disability accrual in relapsing-remitting multiple sclerosis (Sormani *et al.*, 2014) and brain atrophy is a primary outcome measure in phase 2 clinical trials in secondary progressive multiple sclerosis.

A double-blind, controlled, phase 2 clinical trial in long-standing secondary progressive multiple sclerosis (MS-STAT) has shown that high-dose simvastatin reduced the whole brain atrophy rate by 43% with positive effects on frontal lobe function, physician (Expanded Disability Status Scale) and patient-reported (Multiple Sclerosis Impact Scale-29) outcome measures (Chataway *et al.*, 2014; Chan *et al.*, 2017). The MS-STAT trial, like any other randomised-clinical trial, gathered a wealth of information. However, its analysis was limited to separate (univariate), pre-planned outcomes, thereby precluding a mechanistic understanding of the treatment response (Douaud *et al.*, 2013). The pathways via which simvastatin impacts brain atrophy, clinical and cognitive outcomes remain unclear.

Commonly used methods in clinical trials only establish whether one variable affects another but fail to explain *how* such a causal relationship arises. Multivariate mechanistic models can elucidate the *causal* chain of events by simultaneous analysis of multi-modal data that link intermediate variables to outcomes of interest (Bollen and Long, 1992). They have been employed in clinical trials of Alzheimer's disease (Bollen and Long, 1992) and more extensively in social and political sciences (Imai *et al.*, 2011; Kievit *et al.*, 2014). The MS-STAT phase 2 trial offers the unique opportunity to apply multivariate mechanistic models to elucidate the causal sequence of events, in which changes in one variable drive changes in the next one, to explain pathways resulting in the observed treatment effects. Such an understanding is essential to stimulate further mechanistic research in progressive multiple sclerosis to develop new therapies.

A primary characteristic of long-standing progressive MS is a diffuse and extensive neurodegeneration (Frischer *et al.*, 2009; Ransohoff, 2016). By studying a large cohort of secondary progressive patients with a disease duration of 15.6 years, in Chapter 5 I have demonstrated that there is a sequence in which grey matter regions become atrophic, and some regions, such as the basal ganglia, are more atrophic than others (Chapters 3, 4, and 5). The spatiotemporal patterns of brain atrophy in patients with secondary progressive multiple sclerosis and very long disease duration are unknown. The patients recruited in the MS-STAT trial offer the opportunity to address this issue, as their mean disease duration was 21.2 years. Additionally, it is unclear whether these patterns are altered by a medication, such as simvastatin, which significantly reduces the annualised rate of whole brain atrophy.

In this chapter, I re-analysed the MS-STAT trial data and estimated the pathways in which simvastatin *causes* changes in clinical, cognitive and patient-reported outcome measures, either directly or indirectly via brain atrophy. I also investigated whether simvastatin reduces whole brain atrophy rates by a general effect on all brain regions, or focal areas were differentially affected by simvastatin.

6.3. Material and Methods

6.3.1. Participants

This was a *post hoc* study that included all participants of the MS-STAT trial [ClinicalTrials.gov registration number: NCT00647348] performed between 2008-2011 at three research centres and two brain imaging centres in the UK. MS-STAT was a phase-2 double-blind randomised controlled trial whose primary and pre-planned analyses have been reported previously (Chataway *et al.*, 2014; Chan *et al.*, 2017). Briefly, the eligibility criteria were: (i) age between 18-65 years, (ii) Expanded Disability Status Scale of between 4.0 and 6.5, (iii) fulfilling revised 2005 McDonald criteria (see **Chapter 1** for MS diagnostic criteria), and (iv) secondary progressive MS defined by clinically-confirmed disability worsening over the preceding two years. Patients were ineligible if they had corticosteroid treatment or relapse within three months of recruitment, or had received immunomodulatory or immunosuppressive medications within six months of recruitment. Detailed eligibility criteria are available elsewhere (Chataway *et al.*, 2014).

6.3.2. Randomisation

Patients were randomised (1:1) with a centralised server to placebo and high-dose simvastatin (80 mg per day) groups. The randomisation software automatically minimised the following variables between placebo and treatment groups: age (<45 and ≥ 45 years), gender, Expanded Disability Status Scale (4-5.5, and 6.0-6.5), centre (or MRI scanner), and assessing physician. Patients, treating physicians, and outcome assessors were blind to treatment allocation. The treatment allocation was masked to the lead author (Arman Eshaghi) who performed the image analysis. Protocol for compliance with treatment and other details are explained elsewhere (Chataway *et al.*, 2014).

6.3.3. Outcomes

Patients underwent magnetic resonance imaging (MRI), clinical and cognitive assessments at baseline, after one year and two years from the study entry. This study was performed following the Declaration of Helsinki (Association, 2000) and Good Clinical Practice. Berkshire

Research Ethics Committee approved the protocol. Participants gave written informed consent before screening.

6.3.4. Imaging protocol

Patients were scanned at each visit (three visits in total) with 3D T1-weighted, double-echo proton density (PD) and T2-weighted MRI at two imaging centres in the UK with 1.5 Tesla and 3 Tesla scanners. Scanner and MRI protocol remained unique for each participant throughout the trial. “Scanner” was a minimisation variable (as explained above) between treatment and placebo groups. I reported acquisition protocols in the Appendix (**Supplementary Table 6-1**).

6.3.5. Clinical and cognitive outcomes

Patients underwent comprehensive clinical and cognitive assessments. Here, I studied those outcomes that had shown significant (or marginally significant) changes in previous reports, which were the following: The Expanded Disability Status Scale, The Multiple Sclerosis Impact Scale 29v2 (total score and physical subscale)(Hobart *et al.*, 2001), The Wechsler Abbreviated Test of Intelligence (WASI) Block Design (T-score) test (Wechsler, 2011), Paced-auditory serial addition test (PASAT)(Gronwall, 1977), and Frontal Assessment Battery (FAB)(Dubois *et al.*, 2000). Block Design T-score had been calculated against an age-matched reference healthy group from the test manual (Dubois *et al.*, 2000).

6.3.6. Image analysis

I performed image analysis based on the established pipeline for patients with MS (explained in the Methods of Chapter 4). My goals were to extract regional volumes, T2-weighted lesion masks and the whole brain percentage volume change with SIENA (Smith *et al.*, 2001). Briefly, the pipeline included N4-bias field correction of T1-weighted scans to reduce intensity inhomogeneity (Tustison *et al.*, 2010), constructing a symmetric within subject template for unbiased atrophy calculation (Reuter and Fischl, 2011), rigid transformation of T1-weighted, PD, and T2-weighted sequences to this space, automatic longitudinal lesion segmentation of visible T2-weighted lesions with Bayesian Model Selection (BaMoS) (Sudre *et al.*, 2015; Carass *et al.*, 2017), manual editing of these lesion masks and quality assurance with 3D-

Slicer, filling of hypointense lesions in T1-weighted scans (Prados *et al.*, 2016), brain segmentation and parcellation with Geodesic Information Flows (GIF) software (Cardoso *et al.*, 2015). Technical details are explained in detail in **the Supplemental Methods**. **Supplemental Figure 6-1** shows the steps of this pipeline. Outputs of this pipeline were the following: (i) percentage whole brain volume change (SIENA PBVC), (ii) T2-weighted lesion masks, and (iii) regional brain volumes according to Neuromorphometrics' atlas, which is similar to the Desikan-Killiany-Tourville (Klein and Tourville, 2012) atlas available at <http://braincolor.mindboggle.info>, for each region. I summed volumes of the left and right hemispheres.

6.3.7. Statistical analysis

SIENA

I used a linear regression model in which the percentage brain volume change between baseline and two-year follow-up visits was the response variable. This model included treatment allocation as the variable of interest, and the following nuisance variables: age, gender, centre, and Expanded Disability Status Scale. I calculated treatment effect defined as the adjusted difference between percentage whole brain volume change of the two treatment groups, divided by the adjusted percentage whole brain volume change in the placebo group. I set alpha level at 0.05 for all the analyses presented in this work and adjusted univariate analyses for multiple comparisons with the false discovery rate method. I calculated the adjusted difference of the percentage whole brain volume change between placebo and treatment groups and compared it with the original report of this trial (Chataway *et al.*, 2014)

Univariate analysis of T2-weighted lesion load, clinical and cognitive changes

Since the focus of this study was on dynamic changes, I extended the previous analyses of clinical and cognitive outcomes—which were performed as pairwise average comparisons at each baseline and year two visit—to the analyses of rates of change in the two treatment and placebo groups. I aimed to identify variables with a significant difference in their *rates* of changes between the two groups including all the three visits, and to include them in

multivariate mechanistic models (see below). I used univariate linear mixed-effects models in which fixed-effects were time (years from the study entry), and the interaction of time with treatment allocation. I did not include a separate fixed-effects for treatment group; on the assumption that randomisation had worked flawlessly and both groups had similar values for dependent variables at baseline. I confirmed this assumption by comparing baseline values with a separate model including treatment group as an extra fixed-effect and reported the *p*-values. Random effects included time nested in “participant”. To allow for repeated measures, I included random intercept and slope as correlated random effects. In these models, dependent variables were cognitive or clinical outcomes (seven separate models for T2-weighted lesion load, PASAT, Block Design, Expanded Disability Status Scale, FAB, and Multiple Sclerosis Impact Scale 29v2 total and its physical subscale). I included age, gender, and centre as extra fixed-effects (nuisance) variables. I used NLME package (Pinheiro *et al.*, 2017) version 3.1-131 inside R version 3.4.0 (R Core Team, 2014).

Multivariate analysis

I performed multivariate analyses in the following steps:

- (i) Variable selection: to limit the analysis to measures with significant rates of change.
- (ii) Model construction: to formulate mechanistic models as statistical hypotheses.
- (iii) Model selection: to choose the most likely hypothesis.
- (iv) Parameter estimation: to quantify, in the most likely model, pathways between imaging, cognitive, patient-reported, and clinical variables.

Variable selection and model construction

I implemented multivariate analysis with structural equation modelling using Lavaan package version 0.5-23 (Rosseel, 2012) in R. I chose to include those outcomes from the univariate analyses (explained above) that had significant differences in their rate of change between two groups. Since nuisance variables (age, gender, and centre) did not affect the above univariate analyses, I did not include them in multivariate models. I hypothesised seven *a priori* models to explain relationships between these variables according to the literature and my

opinion (Bosma *et al.*, 2015; Larochelle *et al.*, 2016). In these models, the change in brain atrophy (percentage whole brain volume change) was the primary mediator variable that *transmitted* the treatment effects with different pathways in different models to Expanded Disability Status Scale, Block Design and Multiple Sclerosis Impact Scale (e.g., Treatment→SIENA→EDSS or Treatment→SIENA→Block Design in **Figure 6-3**). I only included the physical subtest of the Multiple Sclerosis Impact Scale (instead of the total score), because changes in this subtest drove the change in total score. I also tested treatment effects on the outcomes independent of atrophy, which I refer to as *direct* treatment effects (e.g., Treatment→EDSS or Treatment→Block Design in **Figure 6-3**). Multiple Sclerosis Impact Scale was always assumed to be the last in the cascade of changes because it is a subjective patient-reported questionnaire expected to reflect consequences of clinical and cognitive improvements. **Figure 6-3** shows *a priori* models in order of the increasing complexity from A to G. These models are as follows:

- (A) Full mediation model: treatment effects on disability, cognition and patient-reported outcomes are completely transmitted via brain volume loss. There are no direct treatment effects on disability outcomes.
- (B) Treatment effect is transmitted via brain volume loss to cognitive and clinical outcomes. Clinical and cognitive changes in turn affect the patient-reported outcome. There are no direct effects on clinical or disability outcomes.
- (C) Treatment effect is partially transmitted to slow disability via brain volume loss. Brain volume loss has no effect on the cognitive outcome.
- (D) Treatment effect is partially transmitted to the cognitive outcome, in addition to a direct treatment effect on the cognitive outcome. There is only a direct (but not indirect) treatment effect on the clinical disability.
- (E) Treatment effect is directly and indirectly (via brain atrophy) transmitted to the cognitive and clinical disability outcomes. The patient-reported outcome is only

affected by changes in clinical disability, but not changes in cognition or atrophy.

(F) Treatment effect is directly and indirectly (via brain atrophy) transmitted to the clinical disability and cognitive outcome. The patient-reported outcome is affected by the changes in the clinical disability and brain atrophy but not in cognition.

(G) Partial mediation model: cognition and disability are affected by direct and indirect treatment effects (via brain volume loss). The patient reported outcome is affected by changes in cognitive, disability and brain volume loss.

In these models, I calculated the difference between baseline and second-year values for each variable and divided it by two. I refer to this as the *annualised change* throughout this chapter. In case of SIENA, I entered the percentage whole brain volume change calculated between second-year and baseline scans.

Model selection and parameter estimation

I fitted all the above seven models using full-information maximum likelihood to adjust for missingness, and with the robust standard-errors to account for non-normality (e.g., Expanded Disability Status Scale). I assessed the goodness-of-fit for each model and reported the parameters for the most likely model. To evaluate overall fit of a model I used comparative fit index (CFI; compares the fit of the model with a model with uncorrelated variables; acceptable fit >0.95, good fit >0.97), standardised root mean square residual (SRMSR; square root of the average of the covariance of residuals, good fit <0.08) and root-mean-squared error of approximation (RMSEA; discrepancy between the model and population covariance; good fit <0.06) (Hu and Bentler, 1999). To estimate the relative quality of a model given the data, I calculated information criteria (Akaike information criterion [AIC], and Bayesian information criterion [BIC]) of each model. Interpreting raw AIC and BIC values are cumbersome; therefore, I calculated Akaike and Schwarz weights to represent conditional probability of each model given the data directly (Wagenmakers and Farrell, 2004). To have an unbiased

estimate I calculated fit measures (all those listed above) iteratively on 1000 bootstrap samples and reported the median of bootstrap results.

6.3.8. Regional rates of atrophy

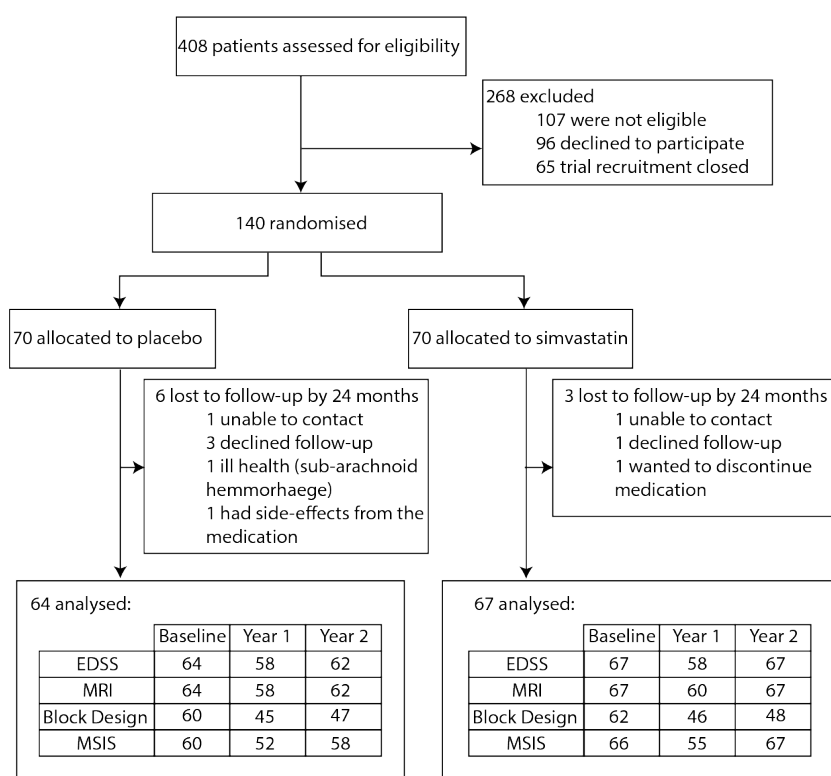
I performed analyses to calculate and compare regional atrophy rates with a univariate mixed-effects model including age, gender, centre, and total intracranial volume to adjust for the head size (Malone *et al.*, 2015). I reported brain regions that had a significant rate of change in the combined treatment and placebo groups as well as separate rates for each treatment group. Further details of statistical modelling are in the Appendix.

6.4. Results

6.4.1. Baseline characteristics

A total of 140 participants were randomised, 9 of whom were lost to follow-up. I, therefore, analysed 131 participants (see **Figure 6-1** for available data at each visit), whom had been scanned with 3 Tesla (total=88, placebo=45, treatment=43) and 1.5 Tesla (total=43, placebo=19, treatment=24) scanners. In the combined analysis of treatment and placebo groups at baseline, the median Expanded Disability Status Scale was 6 (range=4-7), average years (\pm standard deviation) for age was 51.2 (\pm 6.97), for disease duration was 21.17(\pm 8.65), and for progression duration was 7.14 (\pm 5.21). There were twice as many women in this trial as men (ratio=2.04). At baseline, there was no significant difference in age, gender ratio, duration of MS and duration of progression, average years of education, and Expanded Disability Status Scale between treatment and placebo groups. **Table 6-1** shows baseline characteristics of the participants.

Figure 6-1. Trial profile and available data.



This diagram shows the flow of participants from screening to inclusion in the MS-STAT trial.

Available clinical, cognitive, and imaging variables are shown in the table for all the three

visits. EDSS; Expanded Disability Status Scale, MRI; magnetic resonance imaging, MSIS; Multiple Sclerosis Impact Scale.

Table 6-1. Baseline characteristics of the participants.

	Treatment	Placebo	Combined
No. of participants	67	64	131
Average age in years (SD)	51.58 (7.03)	50.9 (6.94)	51.2 (6.97)
Females (males)	46 (21)	42 (22)	88 (43)
Duration of SPMS in years (SD)	7.41 (5.74)	6.87 (4.62)	7.14 (5.21)
Duration of MS in years (SD)	22.32 (8.27)	19.97 (8.94)	21.17 (8.65)
Median EDSS (range)	6 (6-6.5)	6 (4-7)	6 (4-7)
Average years of education (SD)	13.69 (3.07)	13.42 (3.2)	13.56 (3.16)

Abbreviations: SD; standard deviation, SPMS; secondary-progressive multiple sclerosis, MS; multiple sclerosis, EDSS; Kurtzke's Expanded-Disability Status Scale.

6.4.2. The rate of brain atrophy was slower in the treatment group

The average annual percentage whole brain volume change between baseline and second-year in the placebo group was -0.657 (SD=0.62), which was significantly lower (faster atrophy, $P=0.002$) than the treatment group (average = -0.42, SD=0.50). This means that there was a significant reduction of atrophy (Cohen's $d = 0.409$, which shows mild to moderate effect) as measured by SIENA percentage whole brain volume change when comparing simvastatin and placebo groups (**Figure 6-2**). The adjusted difference was 0.245 (95%CI=0.087 to 0.403),

which was indistinguishable from the original report of this trial 0.254 (95% confidence interval [CI]: 0.087 to 0.422).

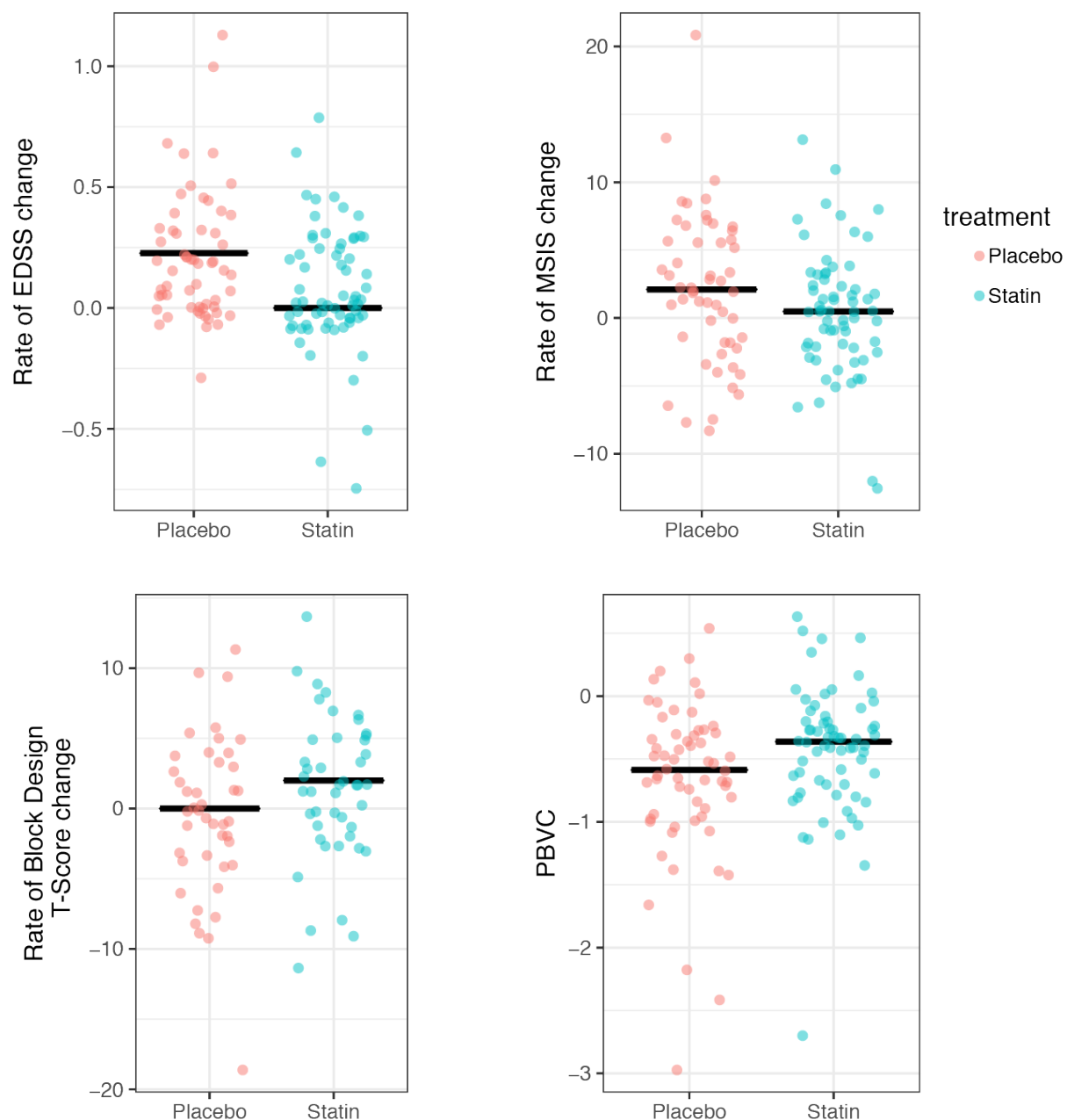
6.4.3. There was no treatment effect on the rate of change in T2-weighted lesion volume

At baseline, lesion volume in the placebo group was 22.14 mL (95%CI: 18.82 to 25.46), which was not different ($p=0.33$) from the treatment group (average=19.3, 95% CI: 13.48 to 25.12). Lesion volumes had a significant rate of change in each group: average [95%CI] for the treatment group was 0.55 ml/year [0.25 to 0.85], and the average for the placebo group was 0.72 ml/year [0.55 to 0.87]). However, rates of change were similar between treatment and placebo groups.

6.4.4. Rate of change was different in some of the clinical and cognitive measures

At baseline, there was no significant difference in the Expanded Disability Status Scale between placebo and treatment groups. However, there was a significant ($P=0.002$) difference in the rate (predicted rate \pm standard-error) of annual Expanded Disability Status Scale worsening between placebo (0.21 ± 0.03) and treatment groups (0.08 ± 0.04). While there was no significant difference in the physical subtest of Multiple Sclerosis Impact Scale 29 at baseline, there was a significant ($P=0.03$) difference in its annual rate of change between the placebo (2.05 ± 0.56) and treatment groups (0.43 ± 0.72). Similarly, the total Multiple Sclerosis Impact Scale-29 score did not show a significant difference at baseline between groups, but its annual rate of change was significantly different ($P=0.03$) between placebo (2.37 ± 0.75) and treatment (0.26 ± 0.97) groups. In Block Design T-score, there was no significant difference at baseline between groups, but there was a significant ($P=0.04$) difference between annual rates of change in placebo (-0.13 ± 0.33) and treatment (0.92 ± 0.45) groups. There were no differences in rates of change between treatment and placebo groups in PASAT and FAB. **Figure 6-2** shows the rate of change between baseline and second-year visits.

Figure 6-2. Outcomes with a significant treatment effect.

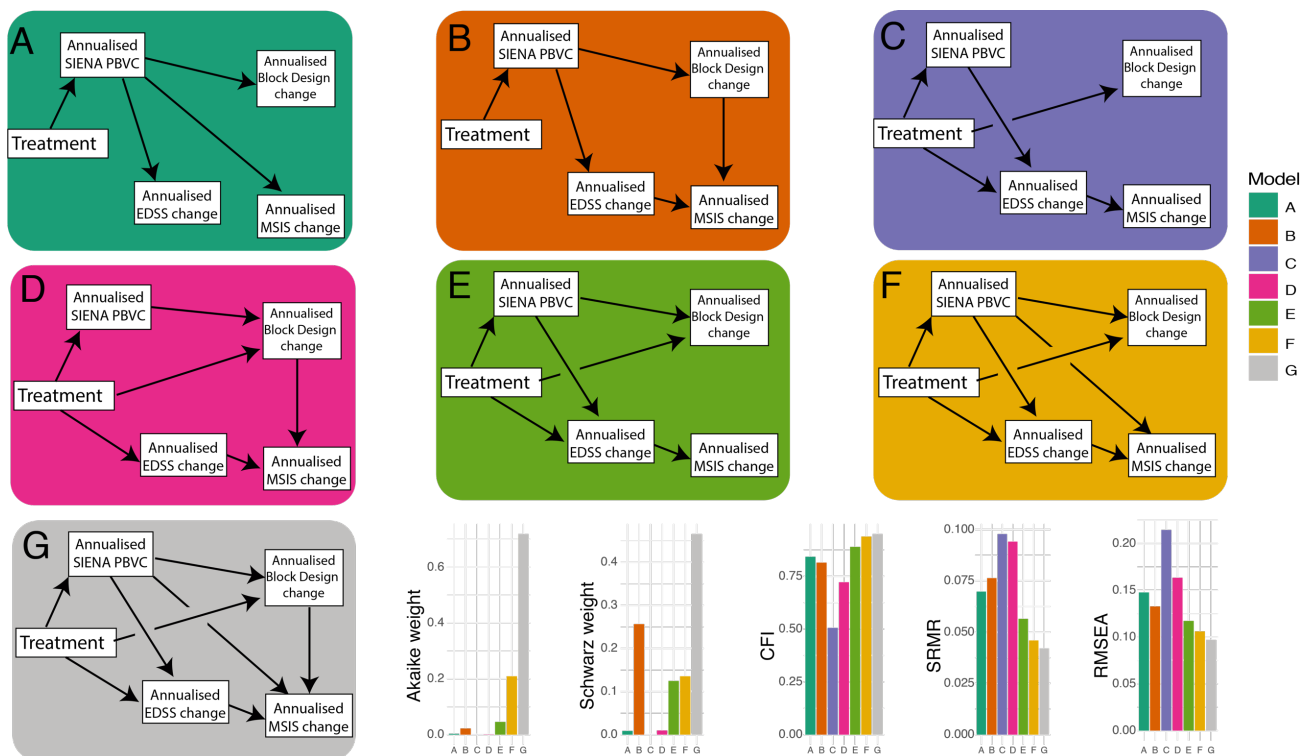


The annual rate of changes for MRI, clinical, cognitive and patient-reported outcomes included in the mechanistic models (EDSS rates are jittered vertically by 0.1 to enable visualising overlapping values). In each of the four plots, horizontal black lines show the medians of the variable shown on y-axes, for placebo (blue) and statin groups (red). EDSS; Expanded Disability Status Scale, MSIS; Multiple Sclerosis Impact Scale, PBVC; percentage brain volume change.

6.4.5. Model G was the most likely among seven a priori models.

Model G showed the best overall fit among the seven models. Bootstrapped fit measures for model G are the following: CFI = 0.95, SRMR = 0.042, $\chi^2 = 4.46$ [p = 0.28], RMSEA = 0.09 [90%Ci=0, 0.18], Akaike weight = 0.71, Schwarz weight = 0.46). Model G was 3.43 times more likely than the second-best model, and 15.71 times more likely than the third best model (E) regarding Kullback–Leibler discrepancy. Model G was 1.8 time ($\frac{\text{Model G Schwarz weight}}{\text{Model F Schwarz weight}} = \frac{0.46}{0.25}$) more likely regarding Schwarz weights than the second-best model (B), and 3.43 times more likely than the third best model (F). **Figure 6-3** shows fit measures for other models.

Figure 6-3. Candidate models and their fit measures.



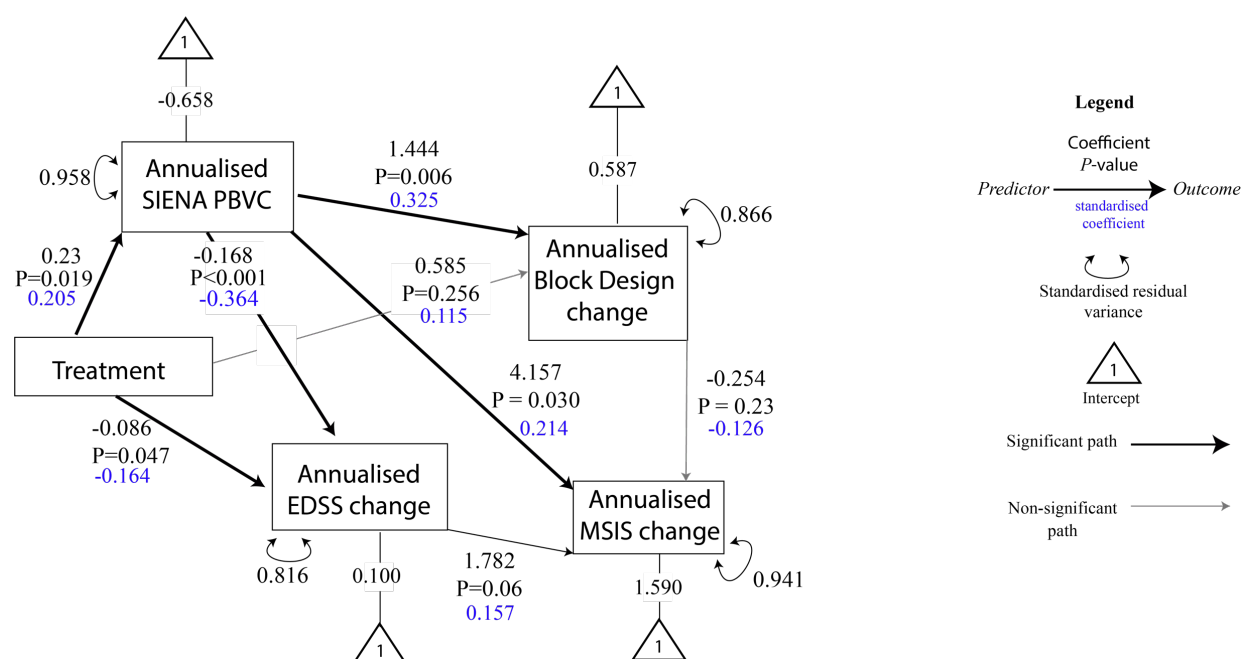
Seven models in order of increasing complexity from A to G are shown. Annualised variables represent rates of change between baseline and second-year follow-up visits. Each rectangle represents a variable. Arrows correspond to regressions between variables, where an arrow starts from a predictor and points to the dependent variable. The bar plots in the bottom row

compare fit-measures that are shown on the y-axis of each bar plots and the seven models on the x-axis. Each colour corresponds to one of the seven models shown in the preceding rows. According to these fit-measures, model G was the most likely model given data, because it had the highest Akaike and Schwarz weights, highest CFI, lowest SRMR, and lowest RMSEA. EDSS; Expanded Disability Status Scale, PBVC; percentage brain volume change, MSIS; Multiple Sclerosis Impact Scale. CFI; confirmatory factor index, SRMR; standardised root mean square residual, RMSEA; root mean squared error of approximation.

Model G showed that both direct (non-standardised estimate=-0.086, standard error [SE]=0.044, $p=0.047$) and indirect effects of simvastatin on the Expanded Disability Status Scale, which were mediated by brain atrophy (beta=-0.039, SE=0.019, $p=0.038$), were significant. Since the estimate of the total effect of simvastatin on the Expanded Disability Status Scale was -0.126 (SE=0.046), this means that 30.9% (-0.039 / -0.126) of the treatment effect on Expanded Disability Status Scale was mediated by percentage whole brain volume change (atrophy), and the remaining 69.1% was *direct* (not explained by atrophy). Although there was a significant association between SIENA percentage whole brain volume change and Block Design T-score (beta=1.44, SE=0.530, $p=0.006$), evidence for the indirect treatment effect (that was mediated by brain atrophy) on Block Design T-score was weak (beta=0.338, SE=0.190, $p=0.075$), which was 33.4% of the total effect of simvastatin on Block Design (beta=1.01, SE=0.52, $p=0.05$). According to the standardised estimates (see **Figure 6-4**) of model G, treatment slowed atrophy to 0.205 SD above the average (holding Expanded Disability Status Scale constant); but only part of this effect was *transmitted* to Expanded Disability Status Scale, such that for each standard deviation reduction in atrophy (faster loss) there was 0.364 SD faster Expanded Disability Status Scale worsening (holding other variables constant). Treatment slowed the rate of Expanded Disability Status Scale worsening to 0.164 SD below the average. Since SD of brain atrophy rate was 0.57, every annual 0.57 reduction in atrophy rate (faster atrophy) corresponded to 0.364 increase in the annual rate of Expanded Disability Status Scale worsening and 0.325 reduction in annual rate of Block

Design scores. According to residual variances (↻ in **Figure 6-4**) model G explains the following shares of variance: 4.2% of SIENA PBVC, 18.4% of annualised Expanded Disability Status Scale change, 5.9% of physical Multiple Sclerosis Impact Scale, and 13.4% of Block Design T-score. **Figure 6-4** shows other estimates.

Figure 6-4. Parameter estimates of the most likely model.



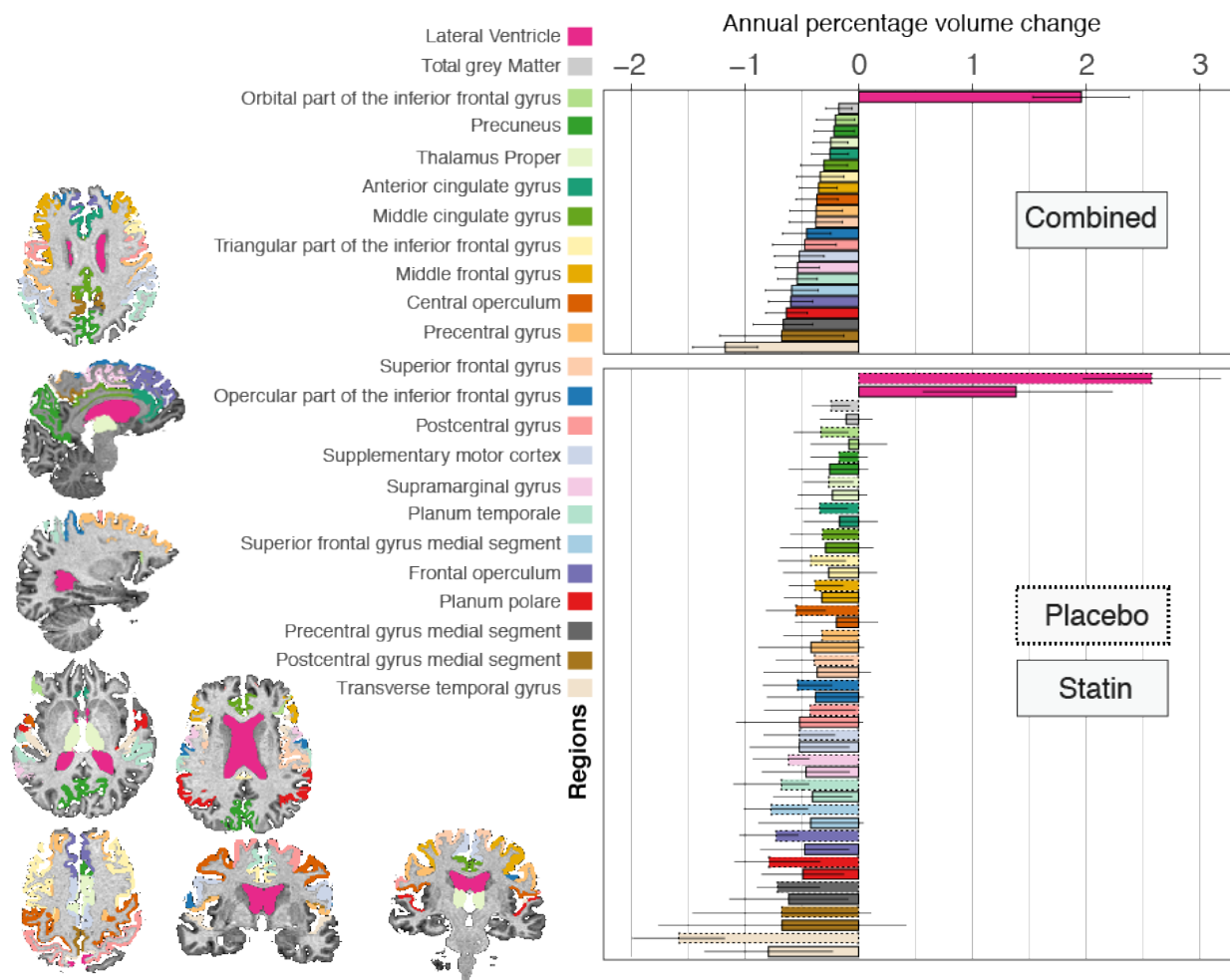
This figure shows the parameter estimates of model G (see also **Figure 6-3**). Each arrow is a regression “path” where the arrow starts from the predictor(s) and points to the dependent variable(s). Significant paths ($p < 0.05$) are shown with bold arrows, while non-significant paths are thinner and grey. Triangles represent intercepts of the model, which are the average changes for the merged placebo and treatment groups for each variable. Black numbers on each arrow represent regression coefficients and their P-values. Blue numbers represent standardised regression coefficient, which can be interpreted as changes in standard deviation (SD). For example, the combined group of placebo and treatment on average has -0.658 of annual PBVC, which is 0.23 higher in the treatment group ($p=0.019$). Treatment increases the PBVC (or slows atrophy) by 0.205 SD. ↻ Indicates standardised residual

variances. For example, the residual variance of annualised EDSS change is 0.816, which means 18.4% (1-0.816) of annualised EDSS change is explained in the model. PBVC; percentage brain volume change, EDSS; Expanded Disability Status Scale, MSIS; Multiple Sclerosis Impact Scale (physical subtest).

6.4.6. Regional analysis

In the analysis of the merged treatment and placebo groups several regions showed significant rate of loss over time, the fastest of which was the lateral ventricle (1.95% annual expansion [1.53%, 2.38%]), and then the transverse temporal gyrus (estimated annual rate = -1.17% [95% CI: -0.88%, -1.46%]). Rates of volume loss in the postcentral and precentral gyri, frontal regions, anterior and middle parts of the cingulate cortex, precuneus, and the thalamus were also significant (see **Figure 6-5** for the full list). When comparing placebo and simvastatin groups, the rates of atrophy were numerically slower in several regions in the simvastatin group (see **Figure 6-5**), however, only the transverse temporal gyrus showed a significant difference ($p=0.002$) in rates of change (estimated annual rate [95% CI] in placebo group = -1.58% [-1.17%, -1.98%]), simvastatin group = -0.79% [-0.22%, -1.35%]) (50% treatment effect).

Figure 6-5. Atrophy rates in areas with significant ongoing change.



This graph shows the adjusted annual rates of volume loss (or expansion for the lateral ventricles) which are calculated from the coefficient of the interaction of time and treatment group in the mixed-effects models constructed separately for each region. Only regions with significant volume change in the combined placebo and treatment analysis are shown (adjusted for multiple comparisons with the false-discovery method). Different colours correspond to different regions that are shown with the same appearance in left on the T1-weighted scan of one of the patients (chosen at random) and, in the right, as bar plots. Two bar plots are shown; the above shows the rate of change in the combined analysis of placebo

and treatment groups on the horizontal axis and different regions on the vertical axis. The lower bar plot shows the rate of change for the same areas for placebo and simvastatin groups separately. This bar plot shows that only the transverse temporal gyrus shows a significant difference in the rate of change when comparing simvastatin and placebo groups. The error bars indicate 95% confidence interval of the rate of change.

6.5. Discussion

In this study, I compared a series of mechanistic hypotheses to investigate *how* a potential neuroprotective drug can influence imaging, clinical, cognitive, and patient-reported outcomes. Simvastatin directly and indirectly, by way of slowing the atrophy rate, slowed the worsening of Expanded Disability Status Scale. It also slowed the worsening of patient-reported outcomes by slowing the rates of atrophy and Expanded Disability Status Scale worsening. Slowing the atrophy rate also led to delaying the deterioration of Block Design test. Moreover, I showed that in long-standing secondary progressive multiple sclerosis, the spatiotemporal pattern of ongoing atrophy was generalised, in that the highest rate of change was seen in the lateral ventricles which is a non-specific and generalised measure of atrophy. The simvastatin effect on brain volume loss was driven by a general reduction in volume loss in multiple regions, but a significant effect was only seen in a region with the highest rate in the grey matter (the transverse temporal gyrus). To the best of my knowledge, this study was the first to explore mechanistic pathways that give rise to clinical and cognitive changes in secondary progressive multiple sclerosis with a potential neuroprotective treatment. The results on the concomitant changes of imaging and clinical outcomes in late-stage secondary progressive MS provide novel insights into *in vivo* pathomechanisms of progressive MS, which have rarely been investigated before. I also reproduced the findings of the original simvastatin study (Chataway *et al.*, 2014), independently, with a novel and blinded image analysis pipeline.

Mechanistic models showed that a reduction in the rate of Expanded Disability Status Scale worsening was partly explained by the treatment effects on brain atrophy, and partly by a separate direct treatment effect. This means that simvastatin reduced the rate of disability worsening via other pathways (not measured directly in the model), which could include vascular, anti-oxidative, diffuse anti-inflammatory effects, or effects on spinal cord atrophy, and are not explained by the treatment effects on brain atrophy alone. A major difference between my study and the previous analyses of MS-STAT, is that I calculated *rates* of change

in imaging and clinical outcomes, rather than average differences between treatment groups at each visit which have been reported before (Chataway *et al.*, 2014; Chan *et al.*, 2017). While Frontal Assessment Battery (FAB) was found to be significantly different between placebo and treatment groups at the last visit in the report of pre-planned statistical analysis of this trial, I did not find a significant difference in its *rate* of change in this study. I, therefore, included Block Design scores in the mechanistic models but not FAB. Block design evaluates the visuospatial memory and depends on fine motor coordination (as it is timed) (Groth-Marnat and Teal, 2000). While there was an association between the rate of brain volume loss and Block Design test, evidence for an indirect treatment effect on this cognitive outcome was weaker than Expanded Disability Status Scale. Mechanistic multivariate models, therefore, have a high potential to quantify and elucidate interrelations between multi-modal measures acquired during a clinical trial.

A novelty of my study was that in contrast to previous studies on convenience samples of patients with (early) progressive multiple sclerosis, I focused on the spatiotemporal pattern of ongoing atrophy in a homogenous population of long-standing secondary progressive multiple sclerosis –a group that has been neglected in multiple sclerosis research (Thompson, 2015). The regional analysis showed that localised atrophy in the temporal lobe, frontal lobe, limbic cortex, and the basal ganglia continues relentlessly in long-standing secondary progressive multiple sclerosis, but the pattern of ongoing atrophy was generalised. Regional susceptibility of neuroanatomical areas to neurodegeneration manifests by faster *percentage* of atrophy rates than that of the entire brain. Annual percentage volume loss is up to 4% in the hippocampus in Alzheimer's disease (Henneman *et al.*, 2009; Josephs *et al.*, 2017), while it is up to 1% for the entire brain. In Huntington's disease, the caudate has 2.1% atrophy rate, while this rate is 1.2% for the entire brain (Tabrizi *et al.*, 2012). In progressive supranuclear palsy, the midbrain shows atrophy rate of 2.2% while this is up to 1.2% in the entire brain (Paviour *et al.*, 2006). In multiple sclerosis, the basal ganglia atrophy rates can be up to 1.5% (see Chapter 4), while the whole brain atrophy is 0.6%. In this study, I found that the highest rate of atrophy across different structures was in the lateral ventricle, which is a non-specific

generalised measure of atrophy. Unlike earlier secondary progressive or primary progressive multiple sclerosis patients, none of the deep grey matter nuclei showed a higher rate than total brain rate (the thalamic atrophy rate was 0.24%), while the whole brain volume loss on average was similar to previous studies (0.65%). My results are in line with pathological observations that generalised neurodegeneration may dominate long-standing secondary progressive multiple sclerosis and become independent of focal inflammation and the lack of effectiveness of anti-inflammatory treatments in late progressive patients (Carassiti *et al.*, n.d.; Hawker *et al.*, 2009), while a more selective pattern of atrophy is seen in earlier multiple sclerosis alongside a more focal inflammation that can respond to immunomodulation (Montalban *et al.*, 2017). These findings support the notion that neuroprotective treatments should become the primary target of drug development in late progressive multiple sclerosis. Although there was a general reduction in several regions in the simvastatin group, only the treatment effect on the transverse temporal gyrus was significant—which also had the highest rate of volume loss in the grey matter. Therefore, a general effect of slowing atrophy rate became detectable in a region with a higher rate. Temporal lobe shows an acceleration in volume loss during secondary progressive multiple sclerosis, which is mostly seen in the insula and deeply located temporal sulci (Haider *et al.*, 2016b). I can speculate that transverse temporal gyrus is spared until later stages of secondary progressive multiple sclerosis, and showed a higher rate after exhaustion of other areas with earlier atrophy.

I used a novel image analysis pipeline alongside SIENA and reproduced the original findings of the MS-STAT trial independently, which was conducted by boundary-shift integral (BSI) and a different segmentation method and registration. The differences between rates of atrophy between placebo and treatment groups were in perfect agreement (average [95%CI] difference between groups in my study: 0.245 [0.087 to 0.403], and in the original report: 0.254 [0.087 to 0.422]). It was initially reported that treatment effect was 43%; however, the treatment effect for this study was 35%. This is a methodological artefact due to a slightly faster average atrophy rates calculated by SIENA (compared to BSI). A previous methodological comparison showed that SIENA produces 20% faster atrophy rates, while

these two methods had an excellent agreement otherwise (Smith *et al.*, 2007). The treatment effect was calculated based on atrophy rates as $\frac{\text{placebo}-\text{treatment}}{\text{placebo}}$. SIENA reduced the treatment effect by increasing the denominator of fraction while the numerator (difference) remained almost the same.

This study had a few limitations. First, information on cardiovascular comorbidities, in which statins have proven effects, was not collected in this trial. Therefore, it remains unclear whether this can influence the treatment response to simvastatin in multiple sclerosis. Secondly, my study is limited by its *post hoc* nature. While pre-planned statistical analyses of clinical trials are the gold-standard to compare treatments, *post hoc* analyses may nevertheless provide information to generate new hypotheses from the wealth of information collected as part of a trial.

In this chapter, I showed that simvastatin mainly affects motor functioning directly, and indirectly by slowing atrophy rates, both of which contribute to improved patient-reported outcomes. A weaker simvastatin effect on visuospatial memory may also exist that is mediated by slowing atrophy rates. I also showed that in long-standing SPMS the *ongoing* atrophy shifts from a localised a more generalised pattern, and that simvastatin had a general effect to reduce the whole brain atrophy rates.

7. Conclusions and future directions

7.1. Novel contributions to the field

In this thesis, I showed that imaging derived measures of brain atrophy have diagnostic and prognostic values and may be used in stratifying patients with multiple sclerosis. The novel contributions of this thesis, which to my knowledge have been reported here for the first time, are as follows:

- 1) Regional grey matter volumes can be used in automatic classification algorithms to distinguish multiple sclerosis from neuromyelitis optica.
- 2) The rate of atrophy in secondary progressive multiple sclerosis accelerates in the cortical regions, most notably in the parietal and temporal lobes.
- 3) Among the grey matter regions, the thalamic volume has the best prognostic value for predicting time to disability progression. This is similar between relapse-onset and primary-progressive multiple sclerosis.
- 4) Longitudinal pattern of regional atrophy progression follows an identifiable sequence which is similar in many regions between relapse-onset and primary progressive multiple sclerosis. This sequence starts from the thalamus, precuneus, posterior cingulate, and brainstem in both relapse-onset and primary progressive multiple sclerosis. The cerebellum showed a differential pattern, which was an early atrophy in relapse-onset multiple sclerosis and a later atrophy in primary progressive multiple sclerosis.
- 5) The sequence of atrophy progression can be used to stratify patients by assigning a stage according to the “number of atrophic brain regions”.
- 6) As reported before, relentless loss of grey matter in secondary progressive multiple sclerosis can be delayed by a potential neuroprotective drug (simvastatin). Here, I showed for the first time that the effects of simvastatin seem to be global, with the highest effect seen in the region with the highest rate of atrophy (transverse temporal gyrus).
- 7) I applied mechanistic multivariate models to test several hypotheses of how simvastatin affects neuroimaging, clinical and cognitive outcomes. I found that the following *causal* chain of events was the most plausible: treatment slows brain atrophy, which in turn slows

the deterioration of cognitive functioning and disability accrual. Reduction in brain atrophy rate, in addition to the reduction of clinical and cognitive deterioration, may cause an improvement in the patient-reported outcome measure.

7.2. Future directions

7.2.1. Large prospective studies

Longitudinal prospective studies of large numbers of patients are lacking in the field of multiple sclerosis. I have therefore retrospectively analysed a large number of patients to calculate rates of change across brain regions (Chapters 4 and 5). Nonetheless, these findings have to be replicated when future prospective studies of multiple sclerosis become available. Prospective longitudinal studies have already been performed in Alzheimer's (Weiner *et al.*, 2017) (Alzheimer's Disease Neuroimaging Initiative) and Parkinson's disease (Parkinson's Progression Markers Initiative) which are success stories for study designs in multiple sclerosis.

7.2.2. Mechanisms of disease progression

An important aim of progressive multiple sclerosis research is the discovery of underlying mechanisms that lead to disability accrual and disease progression (Ontaneda *et al.*, 2017). I expect multivariate mechanistic models (such as the one used in Chapter 6) to become more common in research studies. These models have the potential to formulate complicated scientific hypotheses as tractable statistical models and can, therefore, provide a straightforward answer to research questions of underlying mechanisms. I aim to continue this specific strand of research following my PhD.

7.2.3. Specific and more precise imaging modalities

A major challenge in applying atrophy at the individual patient level was the high variability (e.g., wide confidence intervals of atrophy rates reported in Chapter 4). I, therefore, only showed the importance of imaging-derived atrophy to assess the evolution of atrophy over time in *groups* of patients. I expect in future by understanding the mechanisms of the

physiological variance of structural measures (Nakamura *et al.*, 2015) and improvement in image acquisition techniques (e.g., reducing motion artefacts) to have atrophy measures that can be applied at the individual subject level. Another solution to this problem is the integration of multimodal data to look at the underlying driver of atrophy (which is neurodegeneration). For example, one can look at the rate of neurodegeneration by integrating cerebrospinal fluid markers of neurodegeneration (neurofilament level), with ophthalmological imaging (optical coherence tomography) and magnetic resonance imaging. Other specific neuroimaging modalities such as the positron emission tomography have the potential to address this challenge (Freeman *et al.*, 2015).

7.2.4. Large healthcare records and Big Data analytics

The availability of a large number of digital health data, if tied with available imaging data, can facilitate a better understanding of disease progression in multiple sclerosis (Kalincik *et al.*, 2017; Lorscheider *et al.*, 2017b; Wijnands *et al.*, 2017). These studies will become more informative in future by addition of imaging data to the available large health registries in developed countries (Melão, 2017). Retrospective inspections suffer from inherent limitations. This, in future, can be addressed by real-time monitoring of disability by the analysis of digital sensors that can report several mobility measures using simple digital gadgets (smartphones or watches) (Lopez-Martinez and Picard, 2016; Moon *et al.*, 2017), which may facilitate the understanding of multiple sclerosis disease progression.

Appendix

Chapter 3

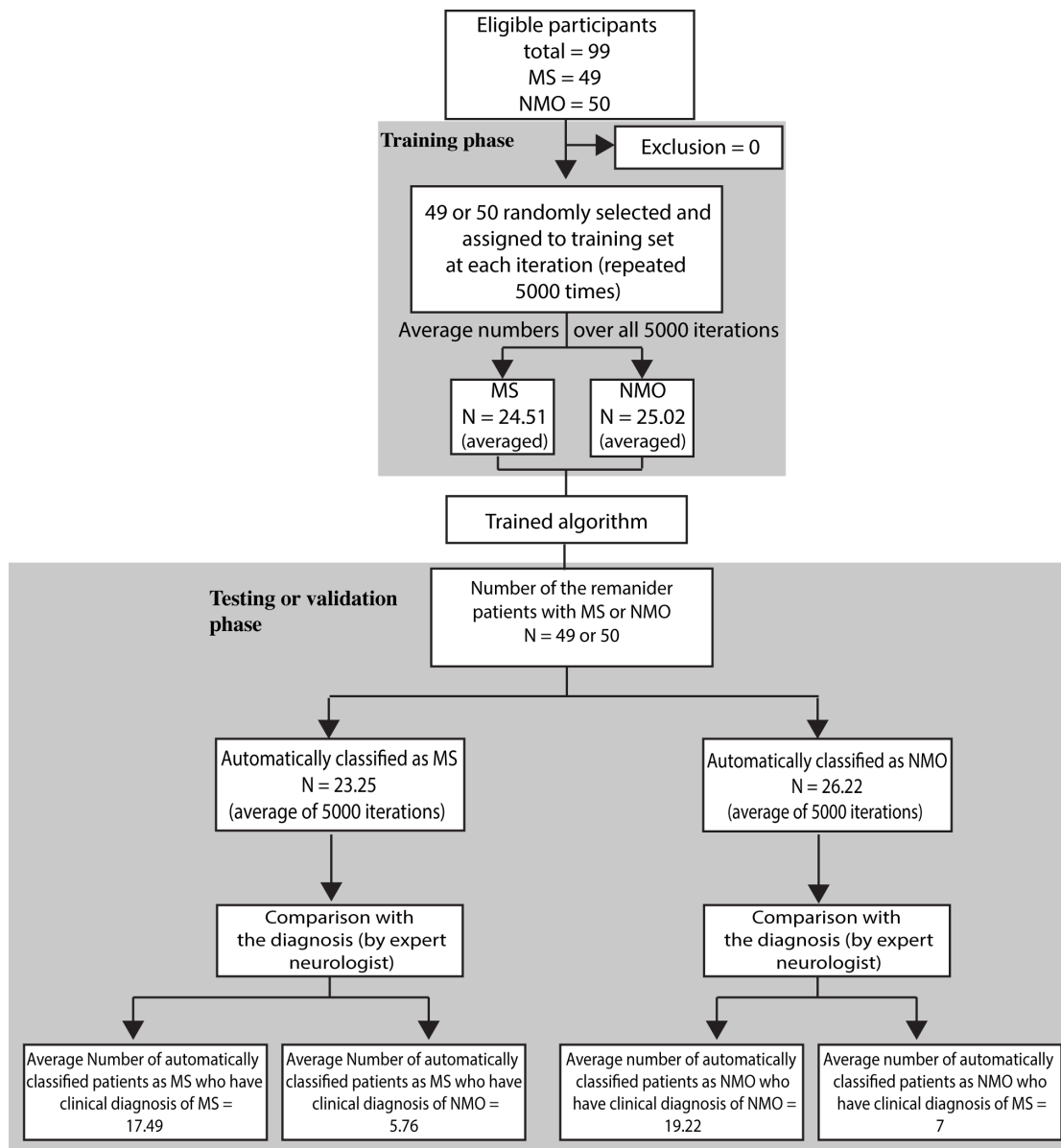
Supplementary Table 3-1. Imaging protocol for each cohort at Tehran and Padua.

Parameter	Tehran	Padua
Field strength	3T	1.5T
Head receiver coil	12 channel	16 channel
Sequence: T1-weighted		
TR/TE	2530 / 3.44 ms	25/4.6 ms
Voxel size	1x1x1 mm	1x1x1 mm
TI	1100 ms	2500 ms
FOV	265 mm	250 mm
Sequence: T2-weighted/FLAIR		
TR/TE	4000/91 ms	1000/120
Slice thickness	3 mm	3 mm
FOV	220 mm	250 mm
Number of slices	42	50

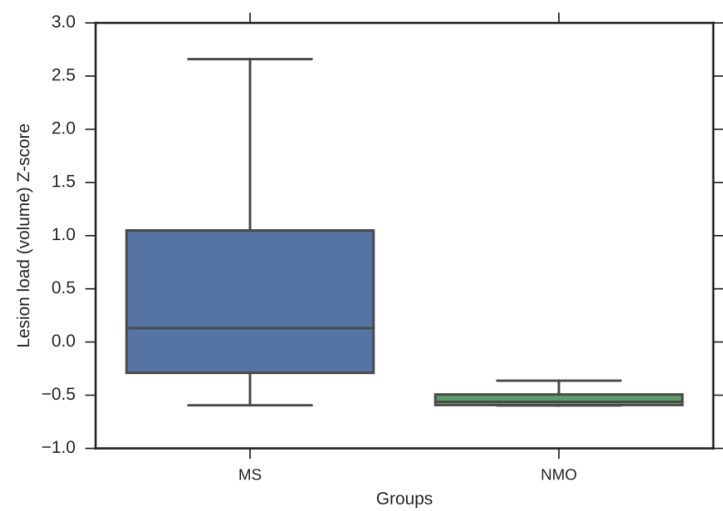
Supplementary Table 3-2. Disease-modifying treatments for each centre.

	Italy		Iran	
	NMO (N=20)	MS (N=24)	NMO (N=30)	MS (N=25)
No treatment	7	3	4	5
Treatments:				
Cyclophosphamide	6	–	1	–
Rituximab	3	–	–	–
Azathioprine	3	–	18	2
Methotrexate	1	–	1	–
Mitoxantrone	–	–	1	1
Interferon beta	–	14	–	17
Mycophenolate mofetil	–	–	5	–
Fingolimod	–	4	–	–
Natalizumab	–	2	–	–
Copaxone	–	1	–	–

Supplementary Figure 0-1. Flow diagram for the diagnostic accuracy of the random-forest classifier in people with multiple sclerosis and neuromyelitis optica.



Supplementary Figure 0-2. Standardised lesion volume in people with multiple sclerosis and neuromyelitis optica.



Chapter 4

Supplementary Table 4-1. T1-weighted MRI parameters and acquisition sites of the included participants.

Centre	London				Milan		Graz	Barcelona		Amsterdam	Rome	Siena	PPMI
Magnetic field	1.5 Tesla			3 Tesla	1.5 Tesla	3 Tesla	3 Tesla	1.5 Tesla	3 Tesla	1.5 Tesla	1.5 Tesla	1.5 Tesla	3 Tesla
Vendor	General Electric Signa	General Electric Signa	General Electric Signa	Philips Achieva	Siemens, Avanto	Philips Intera	Siemens Tim Trio	Siemens Symphony	Siemens Tim Trio	Siemens Vision	Siemens Avanto	Philips Gyroscan	Siemens Tim Trio
Years of recruitment	1999-2006	1998-2005	1998-2010	2011-2016	2007-2014	2008-2015	2006-2013	2013-2016	2010-2016	2004-2007	2012-2016	1999-2013	2010-2015
Included studies	1	1	2	2	1	1	1	1	1	1	1	1	1
Voxel dimension	3D (1.2x1.2x1.5 mm)	3D (1.2x1.2x1.2)	3D (1.2x1.2x1.5)	3D (1x1x1 mm)	3D (1x1x1 mm)	3D (0.89x0.89x1 mm)	3D (1x1x1 mm)	3D (1x1x1 mm)	3D (1x1x1.2 mm)	3D (1x1x1 mm)	3D (1x1x1 mm)	2D (0.97x0.97x3 mm)	3D (1x1x1 mm)
TR	13.3 ms	14.3 ms	29 ms	6.8 ms	2000 ms	25 ms	1900 ms	1980 ms	2300 ms	4000 ms	9000 ms	35 ms	2300
TE	4.2 ms	5.1 ms	15 ms	3.1 ms	3.93 ms	4.6 ms	2.6 ms	3.1 ms	2.98 ms	20 ms	89 ms	10 ms	2.52
Matrix size	256x256	256x256	256x256	256x256	256*224	256*256	176*221	256x256	256x240	180x256	192x256	256x256	176x240
Number of slices	124	156	124	256	208	220	256	176	128	256	160	50	256
Number of participants													
HCs	39	10	0	102	0	0	0	0	0	0	0	23	29
CIS	0	22	60	0	11	0	78	0	76	4	1	1	0
RRMS	33	0	93	46	30	37	77	57	14	141	64	110	0
SPMS	0	6	5	34	11	17	7	5	0	43	6	0	0
PPMS	44	0	0	42	14	0	1	0	0	24	0	0	0

TR, repetition time; TE, echo time; ms, milliseconds; mm, millimetre; HCs, healthy controls; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary-progressive multiple sclerosis; PPMS, primary-progressive multiple sclerosis.

Supplementary Table 4-2. MRI sequence used for lesion delineation.

Centre	London				Milan		Graz	Barcelona		Amsterdam	Rome	Siena	PPMI
Magnetic field	1.5 Tesla			3 Tesla	1.5 Tesla	3 Tesla	3 Tesla	1.5 Tesla	3 Tesla	1.5 Tesla	1.5 Tesla	1.5 Tesla	3 Tesla
Vendor	General Electric Signa	General Electric Signa	General Electric Signa	Philips Achieva	Siemens Avanto	Philips Intera	Siemens Tim Trio	Siemens Symphony	Siemens Tim Trio	Siemens Vision	Siemens Avanto	Philips Gyroscan	Siemens Tim Trio
Sequence	T2-weighted	PD-T2-weighted	PD-T2-weighted	PD-T2-weighted	PD-T2-weighted	PD-T2-weighted	FLAIR	FLAIR	FLAIR	T2-weighted	FLAIR	FLAIR	–
TE	80 ms	17-102 ms	17-102 ms	19-85 ms	28-113 ms	24-120 ms	69 ms	95 ms	93 ms	20 ms	89 ms	150 ms	–
TR	1720 ms	2000 ms	2000 ms	3500 ms	2560 ms	3350 ms	10000 ms	8500 ms	9000 ms	4000 ms	9000 ms	9000 ms	–
TI	–		–	–	–	–	2500 ms	2440 ms	2500 ms	108 ms	2500 ms	2725 ms	–
Matrix size	256x256	256x256	256x 256	240x240	256x256	256x256	192x256	192x256	400x512	256x256	192x256	256x256	–
Slice thickness	5 mm	5 mm	5 mm	3 mm	2.5 mm	3 mm	3 mm	3 mm	3 mm	3 mm	3 mm	3mm	–
Number of slices	28	28	28	50	50	44	44	46	6	44	44	50	–

Supplementary Table 4-3. Rate of annual percentage change in volume.

Mean predicted-slope (95% confidence- interval)	HCs	CIS	RRMS	SPMS	PPMS
Whole brain	-0.05 (0.12, -0.22)	-0.16 (0.01, -0.33)	-0.34 (-0.17, -0.51)	-0.38 (-0.21, -0.55)	-0.46 (-0.29, -0.63)
Cortical GM	-0.34 (-0.1, -0.59)	-0.64 (-0.4, -0.89)	-0.67 (-0.43, -0.91)	-1.11 (-0.87, -1.35)	-0.79 (-0.55, -1.03)
<i>Frontal lobe GM</i>	-0.28 (0.03, -0.59)	-0.62 (-0.3, -0.93)	-0.64 (-0.33, 0.95)	-1.17 (-0.86, -1.48)	-0.86 (-0.55, -1.17)
<i>Temporal lobe GM</i>	-0.52 (-0.23, -0.8)	-0.75 (-0.47, -1.04)	-0.77 (-0.48, -1.05)	-1.21 (-0.93, -1.5)	-0.91 (-0.63, -1.2)
<i>Parietal lobe GM</i>	-0.23 (0.06, -0.52)	-0.63 (-0.34, -0.92)	-0.76 (-0.47, -1.05)	-1.24 (-0.95, -1.52)	-0.9 (-0.61, -1.18)
<i>Occipital lobe GM</i>	-0.43 (-0.09, -0.77)	-0.87 (-0.53, -1.21)	-0.79 (-0.45, -1.13)	-1.08 (-0.73, -1.41)	-0.61 (-0.27, -0.95)
Deep GM	-0.94 (-0.54, -1.33)	-0.88 (-0.49, -1.27)	-1.34 (-0.95, -1.73)	-1.45 (-1.06, -1.84)	-1.66 (-1.27, -2.05)
Cerebellar GM	-0.34 (-0.03, -0.64)	-0.52 (-0.21, -0.82)	-0.58 (-0.28, -0.89)	-0.97 (-0.66, -1.27)	-0.84 (-0.53, -1.14)
Brainstem	-0.31 (0.01, -0.62)	-0.34 (-0.03, -0.66)	-0.67 (-0.35, -0.98)	-0.98 (-0.66, -1.29)	-0.85 (-0.54, -1.17)

White matter rates are not shown because the slopes of change were not significant.

GM; grey matter, HC; healthy control, CIS; clinically isolated syndrome, RRMS; relapsing-remitting multiple sclerosis; SPMS; secondary-progressive multiple sclerosis, PPMS; primary-progressive multiple sclerosis

Supplementary Table 4-4. Pairwise comparison of baseline volumes and rates of change adjusted for multiple comparisons with FDR method.

Regions	HC		CIS		RRMS		SPMS		PPMS	
	Baseline volume	Rate of change	Baseline volume	Rate of change	Baseline volume	Rate of change	Baseline volume	Rate of change	Baseline volume	Rate of change
Whole brain volume										
HC	-	-	ns	ns	<0.001	<0.01	<0.001	<0.01	<0.001	<0.001
CIS			-	-	<0.001	<0.01	<0.001	<0.05	<0.01	<0.01
RRMS					-	-	<0.001	ns	ns	ns
SPMS							-	-	<0.05	ns
PPMS									-	-
GM of the whole cortex										
HC	-	-	<0.05	ns	<0.001	<0.05	<0.001	<0.01	<0.001	<0.01
CIS			-	-	<0.001	ns	<0.001	<0.05	<0.05	ns
RRMS					-	-	<0.001	ns	ns	ns
SPMS							-	-	ns	ns
PPMS									-	-
Frontal lobe GM										
HC	-	-	ns	ns	<0.01	<0.05	<0.001	<0.01	<0.01	<0.01
CIS			-	-	<0.001	ns	<0.001	ns	<0.01	ns
RRMS					-	-	<0.001	ns	ns	ns
SPMS							-	-	ns	ns
PPMS									-	-
Temporal lobe GM										
HC	-	-	<0.01	ns	<0.001	<0.05	<0.001	<0.001	<0.001	<0.01
CIS			-	-	<0.001	ns	<0.001	<0.05	<0.05	ns
RRMS					-	-	<0.001	<0.05	ns	ns
SPMS							-	-	ns	ns
PPMS									-	-
Parietal lobe GM										
HC	-	-	<0.05	ns	<0.001	<0.01	<0.001	<0.01	<0.001	<0.01
CIS			-	-	<0.001	ns	<0.001	<0.05	<0.01	ns
RRMS					-	-	<0.001	ns	ns	ns
SPMS							-	-	<0.05	ns
PPMS									-	-
Occipital lobe GM										
HC	-	-	<0.001	ns	<0.001	<0.05	<0.001	<0.01	<0.001	ns
CIS			-	-	<0.01	ns	<0.01	ns	ns	ns
RRMS					-	-	ns	ns	ns	ns
SPMS							-	-	ns	ns
PPMS									-	-
Deep GM										
HC	-	-	<0.001	ns	<0.001	<0.01	<0.001	<0.01	<0.001	<0.01
CIS			-	-	<0.001	<0.01	<0.001	<0.01	<0.05	<0.01
RRMS					-	-	<0.001	ns	ns	ns
SPMS							-	-	<0.001	ns
PPMS									-	-
Cerebellar GM										
HC	-	-	<0.05	ns	<0.001	ns	<0.001	<0.05	<0.01	<0.05
CIS			-	-	<0.001	ns	<0.001	<0.05	ns	ns

RRMS					-	-	<0.001	ns	ns	ns
SPMS							-	-	ns	ns
PPMS									-	-
Brainstem										
HC	-	-	<0.001	ns	<0.001	<0.05	<0.001	<0.001	<0.001	<0.01
CIS			-	-	ns	<0.05	<0.001	<0.01	<0.01	ns
RRMS					-	-	<0.01	<0.05	<0.05	ns
SPMS							-	-	ns	ns
PPMS									-	-

Each cell shows the corrected *p*-values of the comparison between the groups in the corresponding column and row.

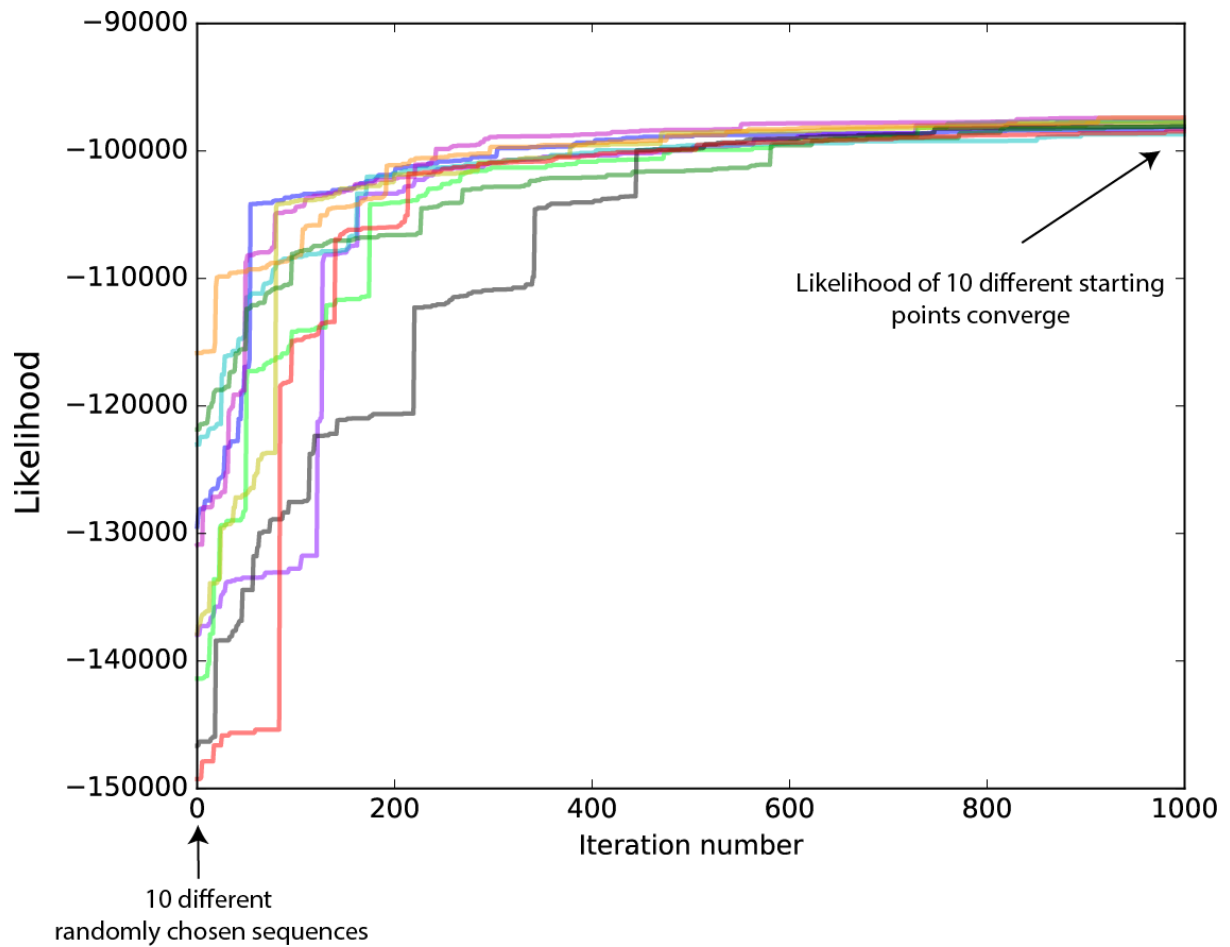
Abbreviations: HC, healthy control; CIS, clinically isolated syndrome; RRMS, relapsing-remitting multiple sclerosis; SPMS, secondary-progressive multiple sclerosis; PPMS, primary-progressive multiple sclerosis; ns, non-significant.

Chapter 5

Supplementary Table 1. MRI protocols for each participating centre.

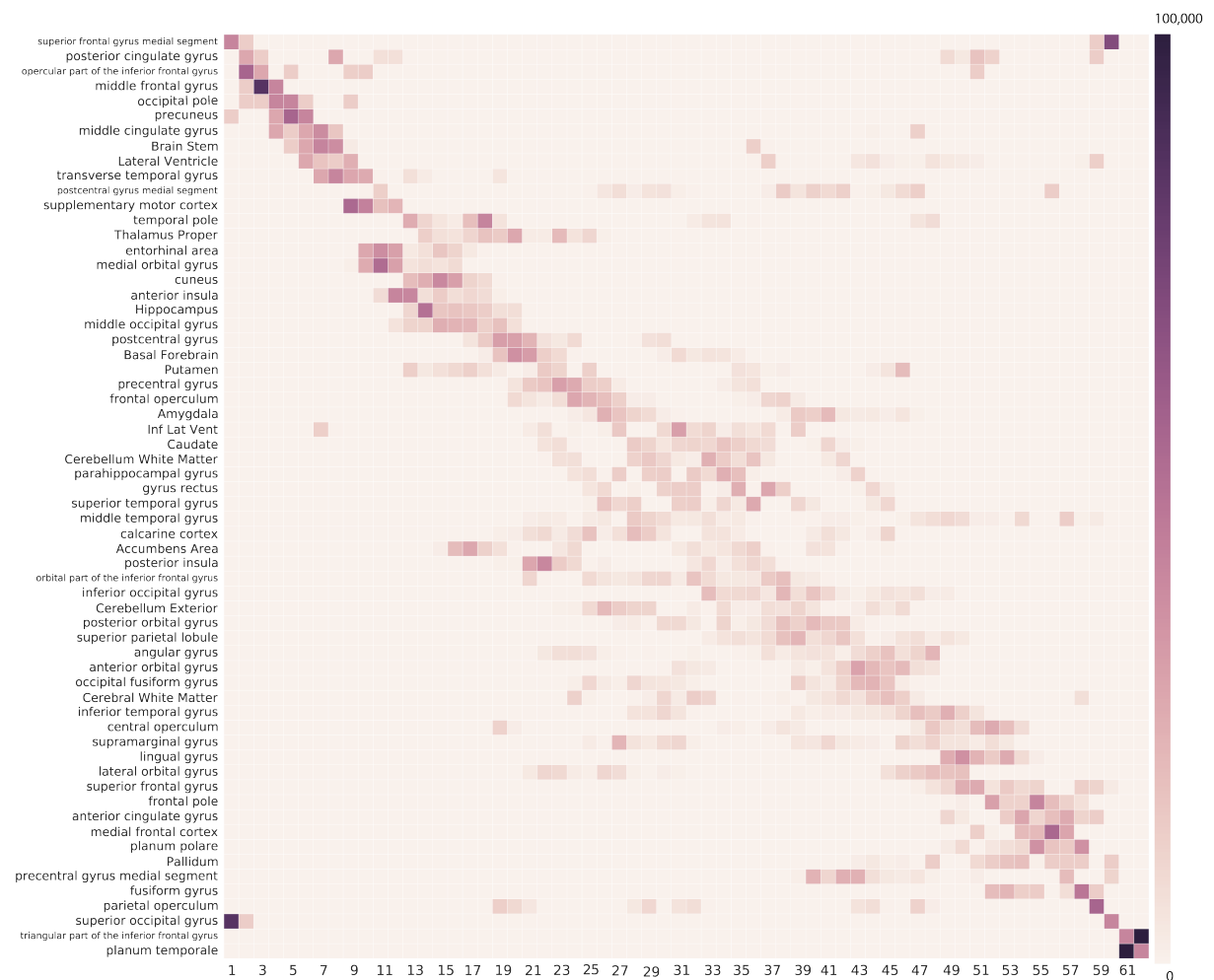
	T1-weighted MRI							Sequence for lesion delineation					
	Magnetic field	Vendor	Voxel dimension	TR	TE	Matrix size	Slices	MR Sequence	TE	TR	TI	Matrix size	Slice thickness
London	1.5T	General Electric Signa	3D (1.2x1.2x1.5 mm)	13.3 ms	4.2 ms	256x256	124	T2	80 ms	1720 ms	–	256x256	5 mm
	1.5T	General Electric Signa	3D (1.2x1.2x1.2)	14.3 ms	5.1 ms	256x256	156	PD-T2	17-102 ms	2000 ms		256x256	5 mm
	1.5T	General Electric Signa	3D (1.2x1.2x1.5)	29 ms	15 ms	256x256	124	PD-T2	17-102 ms	2000 ms	–	256x 256	5 mm
	3T	Philips Achieva	3D (1x1x1 mm)	6.8 ms	3.1 ms	256x256	256	PD-T2	19-85 ms	3500 ms	–	240x240	3 mm
Milan	1.5T	Siemen Avanto	3D (1x1x1 mm)	2000 ms	3.93 ms	256*224	208	PD-T2	28-113 ms	2560 ms	–	256x256	2.5 mm
	3T	Philips Intera	3D (0.89x0.89x1 mm)	25 ms	4.6 ms	256*256	220	PD-T2	24-120 ms	3350 ms	–	256x256	3 mm
Graz	3T	Siemens Tim Trio	3D (1x1x1 mm)	1900 ms	2.6 ms	176*221	256	FLAIR	69 ms	10000 ms	2500 ms	192x256	3 mm
Barcelona	1.5T	Siemens Symphony	3D (1x1x1 mm)	1980 ms	3.1 ms	256x256	176	FLAIR	95 ms	8500 ms	2440 ms	192x256	3 mm
	3T	Siemens Tim Trio	3D (1x1x1.2 mm)	2300 ms	2.98 ms	256x240	128	FLAIR	93 ms	9000 ms	2500 ms	400x512	3 mm
Amsterdam	1.5T	Siemens Vision	3D (1x1x1 mm)	4000 ms	20 ms	180x256	256	T2	20 ms	4000 ms	108 ms	256x256	3 mm
Rome	1.5T	Siemens Avanto	3D (1x1x1 mm)	9000 ms	89 ms	192x256	160	FLAIR	89 ms	9000 ms	2500 ms	192x256	3 mm
Siena	1.5T	Philips Gyroscan	2D (0.97x0.97x3mm)	35 ms	10 ms	256x256	50	FLAIR	150 ms	9000 ms	2725 ms	256x256	3mm
PPMI	3T	Siemens Tim Trio	3D (1x1x1mm)	2300 ms	2.52 ms	176x240	256	–	–	–	–	–	–

Supplementary Figure 0-3. Greedy ascent search.

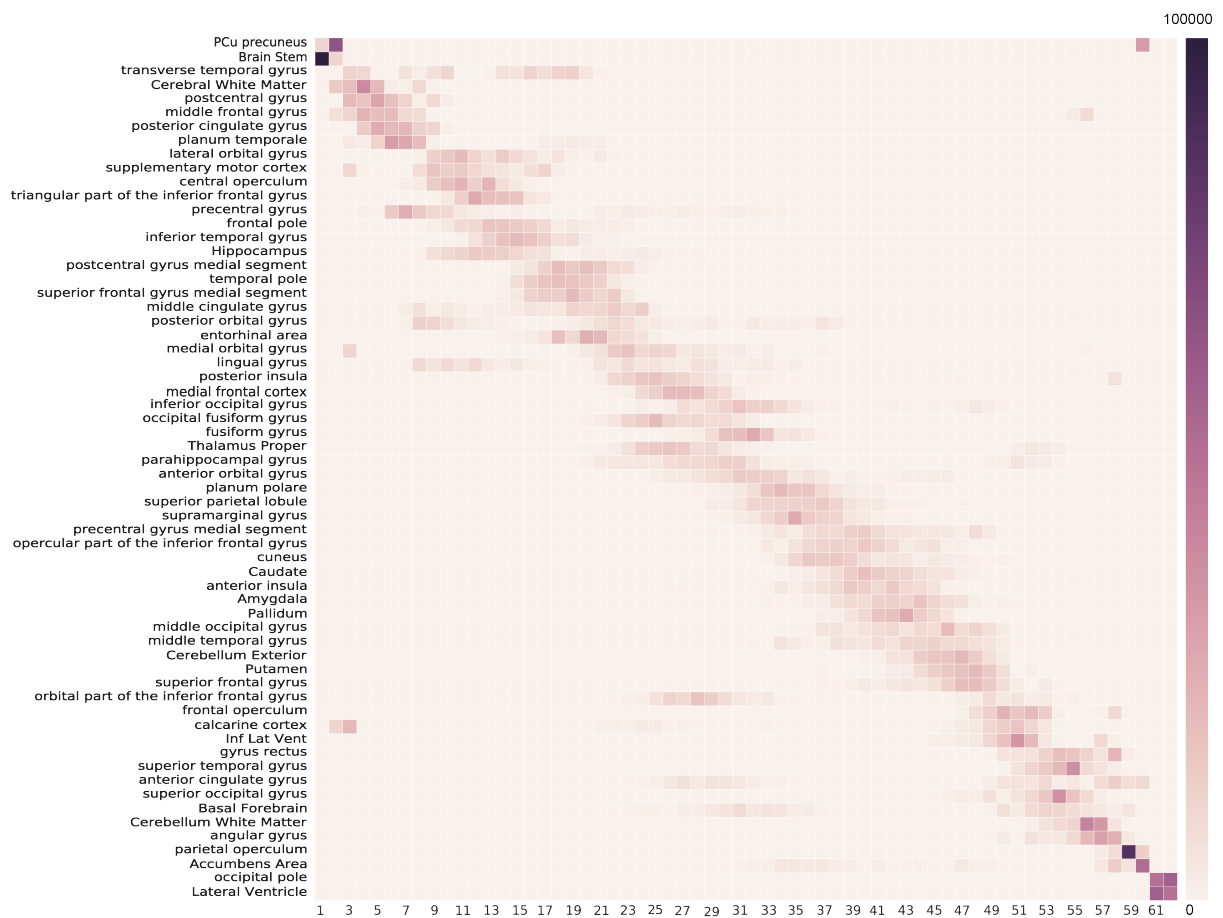


The most likely sequence of atrophy progression in relation to 10 randomly chosen initial sequences. The y-axis shows the data likelihood (calculated from Equation 1). The x-axis shows the number of iterations at which two events are randomly swapped in search for a higher sequence likelihood. This procedure was repeated during each cross-validation (10 times).

Supplementary Figure 0-4. Positional variance diagram for CIS/relapse-onset MS based on all brain regions.



Supplementary Figure 0-5. Positional variance diagram for PPMS based on all brain regions.



Chapter 6

Supplemental methods

Image analysis

N4-bias field correction

I used ANTs software version 2.2 (Tustison *et al.*, 2010) to correct for the scanner-field inhomogeneity in T1-weighted scans. I used Montreal Neurological Institute intracranial mask (Boyes *et al.*, 2008) transferred with diffeomorphic registration (Avants *et al.*, 2008) to the native space to limit the correction to the cranium.

Symmetric within-subject template construction

I constructed an isotropic symmetric template per subject using available time-points with iterative rigid registration (Reuter and Fischl, 2011; Leung *et al.*, 2012). This step is necessary to avoid bias towards a time-point (e.g., baseline) since it distributes interpolation and segmentation errors across time-points for an unbiased atrophy calculation (Fox *et al.*, 2011).

Symmetric transformation

I transferred T1-weighted, PD and T2-weighted scans to the within-subject template by applying the symmetric transformation matrix. I reconstructed scans with B-spline interpolation to minimise blurring artefacts.

Automatic lesion segmentation

I used Bayesian Model Selection (BaMoS) to segment white matter lesions longitudinally and produce lesion masks (Sudre *et al.*, 2017). BaMoS is a multimodal method that integrates PD, T2-weighted, and T1-weighted segmentations to provide lesion masks.

Manual editing

I used 3D-Slicer (<https://www.slicer.org>) version 4.6 to manually edit lesion masks acquired from BaMoS.

White matter segmentation

I used Geodesic Information Flows (GIF) version 3.0 to segment T1-weighted scans and calculate (normal-appearing) white matter masks. This mask enables filling hypointense white matter lesions while avoiding any change in ventricular sizes (Prados *et al.*, 2016).

T1-weighted hypointense lesion filling

I used a longitudinal patch-based method to fill hypointense lesions on T1-weighted scans (Prados *et al.*, 2016). I used white matter mask from the previous step as a reference to fill hypointense lesions. This step minimises erroneous segmentation of hypointense-lesions as grey matter and increases the precision of atrophy estimates as explained elsewhere (Prados *et al.*, 2016).

Brain segmentation and parcellation

I used GIF to segment lesion-filled T1-weighted scans into grey matter, white matter, and CSF and to parcellate the brain to ~120 regions according to the Desikan-Killiany-Tourville protocol (<http://braincolor.mindboggle.info/index.html>). I calculated the volume of each parcellated region by multiplying segmentation probability maps with the voxel volume.

To calculate whole brain percentage atrophy I used SIENA (part of FSL version 5.0) (Smith *et al.*, 2001). SIENA estimates the rate of atrophy by measuring the shift of brain edge over two separate time-points. To have consistent results between regional and global atrophy that were not limited by the differences in segmentation methods,

I used GIF masks within SIENA instead of BET (Smith, 2002) and FAST (Zhang *et al.*, 2001).

Statistical analysis

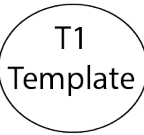
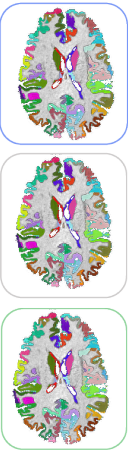
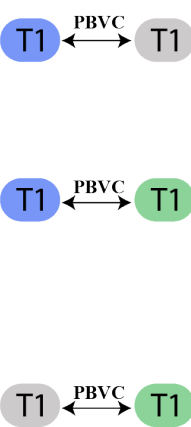
Regional analysis

To explore regional treatment effects, and primary drivers of the total brain atrophy I summed respective regions from left and right hemispheres and constructed linear mixed-effects models for each area (~60 models), where the volume of a given area was the dependent variable. Independent variables (fixed effects and random effects) were similar to the models used for cognitive and clinical outcomes with an additional variable for total intracranial volume to adjust for the head size and scanner (1.5 Tesla or 3 Tesla). First, I extracted rates of atrophy for those regions that had a significant rate of change (significant slope), after adjustment for multiple comparisons with the false-discovery rate (Benjamini and Hochberg, 1995). With a similar model, I calculated the rate of change for treatment and placebo groups.

Supplementary table 6-1. MRI protocol.

Sequence: T1-weighted		
Centre	One	Two
Vendor	General electric Signa	Siemens
Magnetic field	3 Tesla	1.5 Tesla
Voxel dimension	3D (0.93x0.93x1.1 mm)	3D (1.25x1.25x1.2)
Repetition time	7.808 ms	2400 ms
(TR)		
Echo time (TE)	3.004 ms	3.45
Acquisition matrix	256x256	192x192
Inversion time	450 ms	1000 ms
Flip angle	20	8
Field of view		192x192
Number of slices	170	160
Sequence: dual echo proton density and T2-weighted		
Centre	One	Two
Voxel dimension	0.97x0.97x3.0	0.48x0.48x3.0
Acquisition type	2D	2D
Repetition time (TR)	2600 ms	4220 ms
Spacing between slices	3 mm	3 mm
Echo train length	10	5
Acquisition matrix	256x256	424x512
Flip angle	90	150
Number of slices	46	46

Supplementary Figure 0-6. Image analysis pipeline.

Raw data	1- N4 bias-field correction	2- Creating symmetric within-subject template	3- Symmetric transformation of time-points to the template	4- Longitudinal lesion segmentation (BaMoS)	5- Manual editing of white-matter lesion masks
Baseline T1 Baseline T2 Baseline PD Year-1 T1 Year-1 T2 Year-1 PD Year-2 T1 Year-2 T2 Year-2 PD	Baseline T1 Year-1 T1 Year-2 T1		Baseline T1 Baseline T2 Baseline PD Year-1 T1 Year-1 T2 Year-1 PD Year-2 T1 Year-2 T2 Year-2 PD	Baseline T1 Baseline T2 Baseline PD Year-1 T1 Year-1 T2 Year-1 PD Year-2 T1 Year-2 T2 Year-2 PD	Baseline lesion mask Year-1 lesion mask Year-2 lesion mask
6- Normal-appearing white matter segmentation	7- Multi-time point lesion filling	8- Segmentation and parcellation	9- SIENA with GIF segmentations in the symmetric-space	10- Mixed-effects modelling of regional and whole-brain measures	
T1 White-matter mask Lesion mask T1 White-matter mask Lesion mask T1 White-matter mask Lesion mask	lesion-filled T1 lesion-filled T1 lesion-filled T1			<div>Whole brain</div> <p>Outcome variable: SIENA PBVC</p> <p>Time variable: Interval between scans</p>	
				<div>Regional</div> <p>Outcome variable: volume</p> <p>Time variable: Interval from study entry</p>	

References

- Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* 2014; 8: 14.
- Amato MP, Hakiki B, Goretti B, Rossi F, Stromillo ML, Giorgio A, et al. Association of MRI metrics and cognitive impairment in radiologically isolated syndromes. *Neurology* 2012; 78: 309–314.
- Amiry-Moghaddam M, Ottersen OP. The molecular basis of water transport in the brain. *Nat Rev Neurosci* 2003; 4: 991–1001.
- Ashburner J, Friston KJ. Voxel-Based Morphometry—The Methods. *NeuroImage* 2000; 11: 805–821.
- Ashburner J, Friston KJ. Why Voxel-Based Morphometry Should Be Used. *NeuroImage* 2001; 14: 1238–1243.
- Ashburner J, Friston KJ. Unified segmentation. *Neuroimage* 2005; 26: 839–51.
- Association WM. Declaration of Helsinki, ethical principles for medical research involving human subjects. 52 Nd WMA Gen. Assem. Edinb. Scotl. 2000
- Audoin B, Zaaraoui W, Reuter F, Rico A, Malikova I, Confort-Gouny S, et al. Atrophy mainly affects the limbic system and the deep grey matter at the first stage of multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 2010; 81: 690–695.
- Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal* 2008; 12: 26–41.
- Avants BB, Yushkevich P, Pluta J, Minkoff D, Korczykowski M, Detre J, et al. The optimal template effect in hippocampus studies of diseased populations. *Neuroimage* 2010; 49: 2457–66.
- Baranzini SE, Oksenberg JR. The Genetics of Multiple Sclerosis: From 0 to 200 in 50 Years [Internet]. *Trends Genet.* 2017 Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0168952517301634>
- Barkhof F. The clinico-radiological paradox in multiple sclerosis revisited. *Curr. Opin. Neurol.* 2002; 15: 239–245.
- Barkhof F. Brain atrophy measurements should be used to guide therapy monitoring in MS - NO. *Mult. Scler. Houndmills Basingstoke Engl.* 2016
- Battaglini M, Jenkinson M, De Stefano N. Evaluating and reducing the impact of white matter lesions on brain volume measurements. *Hum. Brain Mapp.* 2012; 33: 2062–2071.
- Bendfeldt K, Hofstetter L, Kuster P, Traud S, Mueller-Lenke N, Naegelin Y, et al. Longitudinal gray matter changes in multiple sclerosis--differential scanner and overall disease-related effects. *Hum. Brain Mapp.* 2012; 33: 1225–1245.
- Bendfeldt K, Klöppel S, Nichols TE, Smieskova R, Kuster P, Traud S, et al. Multivariate pattern classification of gray matter pathology in multiple sclerosis. *NeuroImage* 2012; 60: 400–408.

Bendfeldt K, Smieskova R, Koutsouleris N, Klöppel S, Schmidt A, Walter A, et al. Classifying individuals at high-risk for psychosis based on functional brain activity during working memory processing. *NeuroImage Clin.* 2015; 9: 555–563.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* 1995; 57: 289–300.

Bermel RA, Bakshi R. The measurement and clinical relevance of brain atrophy in multiple sclerosis. *Lancet Neurol.* 2006; 5: 158–170.

Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models. *NeuroImage* 2013; 66: 249–260.

Biberacher V, Schmidt P, Keshavan A, Boucard CC, Righart R, Sämann P, et al. Intra- and interscanner variability of magnetic resonance imaging based volumetry in multiple sclerosis. *NeuroImage* 2016

Blaiotta C, Freund P, Cardoso MJ, Ashburner J. Generative diffeomorphic modelling of large MRI data sets for probabilistic template construction [Internet]. *NeuroImage* Available from: <https://www.sciencedirect.com/science/article/pii/S1053811917308947>

Blanc F, Noblet V, Jung B, Rousseau F, Renard F, Bourre B, et al. White matter atrophy and cognitive dysfunctions in neuromyelitis optica. *PLoS One* 2012; 7: e33878.

Bocchetta M, Cardoso MJ, Cash DM, Ourselin S, Warren JD, Rohrer JD. Patterns of regional cerebellar atrophy in genetic frontotemporal dementia. *NeuroImage Clin.* 2016; 11: 287–290.

Bodini B, Chard D, Altmann DR, Tozer D, Miller DH, Thompson AJ, et al. White and gray matter damage in primary progressive MS: The chicken or the egg? *Neurology* 2016; 86: 170–176.

Bollen KA, Long JS. Tests for Structural Equation Models: Introduction. *Sociol. Methods Res.* 1992; 21: 123–131.

Bonavita S, Gallo A, Sacco R, Corte MD, Bisecco A, Docimo R, et al. Distributed changes in default-mode resting-state connectivity in multiple sclerosis. *Mult. Scler. Houndmills Basingstoke Engl.* 2011; 17: 411–422.

Bosma LV a. E, Sonder JM, Kragt JJ, Polman CH, Uitdehaag BMJ. Detecting clinically-relevant changes in progressive multiple sclerosis. *Mult. Scler. Houndmills Basingstoke Engl.* 2015; 21: 171–179.

Boyes RG, Gunter JL, Frost C, Janke AL, Yeatman T, Hill DLG, et al. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *NeuroImage* 2008; 39: 1752–1762.

Breiman L. Random Forests. *Mach. Learn.* 2001; 45: 5–32.

Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and Regression Trees. New Ed edition. New York: Chapman and Hall/CRC; 1984.

- Brown JWL, Pardini M, Brownlee WJ, Fernando K, Samson RS, Prados Carrasco F, et al. An abnormal periventricular magnetization transfer ratio gradient occurs early in multiple sclerosis. *Brain* 2017; 140: 387–398.
- Browne P, Chandraratna D, Angood C, Tremlett H, Baker C, Taylor BV, et al. Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity. *Neurology* 2014; 83: 1022–1024.
- Brownlee WJ, Hardy TA, Fazekas F, Miller DH. Diagnosis of multiple sclerosis: progress and challenges. *The Lancet* 2017; 389: 1336–1346.
- Calabrese M, Magliozzi R, Ciccarelli O, Geurts JJG, Reynolds R, Martin R. Exploring the origins of grey matter damage in multiple sclerosis. *Nat. Rev. Neurosci.* 2015; 16: 147–158.
- Calabrese M, Oh MS, Favaretto A, Rinaldi F, Poretto V, Alessio S, et al. No MRI evidence of cortical lesions in neuromyelitis optica. *Neurology* 2012; 79: 1671–6.
- Calabrese M, Reynolds R, Magliozzi R, Castellaro M, Morra A, Scalfari A, et al. Regional distribution and evolution of gray matter damage in different populations of multiple sclerosis patients. *PloS One* 2015; 10: e0135428.
- Cannon TD, Sun F, McEwen SJ, Papademetris X, He G, van Erp TGM, et al. Reliability of neuroanatomical measurements in a multisite longitudinal study of youth at risk for psychosis. *Hum. Brain Mapp.* 2014; 35: 2424–2434.
- Carass A, Roy S, Jog A, Cuzzocreo JL, Magrath E, Gherman A, et al. Longitudinal multiple sclerosis lesion segmentation: Resource and challenge. *NeuroImage* 2017; 148: 77–102.
- Carassiti D, Altmann DR, Petrova N, Pakkenberg B, Scaravilli F, Schmierer K. Neuronal loss, demyelination and volume change in the multiple sclerosis neocortex. *Neuropathol. Appl. Neurobiol.*: n/a-n/a.
- Cardoso MJ, Modat M, Wolz R, Melbourne A, Cash D, Rueckert D, et al. Geodesic information flows: spatially-variant graphs and their application to segmentation and fusion. *IEEE Trans. Med. Imaging* 2015; 34: 1976–1988.
- Cash DM, Frost C, Iheme LO, Ünay D, Kandemir M, Fripp J, et al. Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge. *NeuroImage* 2015; 123: 149–164.
- Cawley N, Solanky BS, Muhlert N, Tur C, Edden RAE, Wheeler-Kingshott CAM, et al. Reduced gamma-aminobutyric acid concentration is associated with physical disability in progressive multiple sclerosis. *Brain J. Neurol.* 2015; 138: 2584–2595.
- Ceccarelli A, Rocca MA, Pagani E, Colombo B, Martinelli V, Comi G, et al. A voxel-based morphometry study of grey matter loss in MS patients with different clinical phenotypes. *NeuroImage* 2008; 42: 315–322.
- Chan D, Binks S, Nicholas JM, Frost C, Cardoso MJ, Ourselin S, et al. Effect of high-dose simvastatin on cognitive, neuropsychiatric, and health-related quality-of-life measures in secondary progressive multiple sclerosis: secondary analyses from the MS-STAT randomised, placebo-controlled trial. *Lancet Neurol.* 2017; 16: 591–600.

Chanson JB, Lamy J, Rousseau F, Blanc F, Collongues N, Fleury M, et al. White matter volume is decreased in the brain of patients with neuromyelitis optica. *Eur J Neurol* 2013; 20: 361–7.

Chataway J, Schuerer N, Alsanousi A, Chan D, MacManus D, Hunter K, et al. Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial. *Lancet Lond. Engl.* 2014; 383: 2213–2221.

Chen C, Liaw A, Breiman L. Using random forest to learn imbalanced data. Univ. Calif. Berkeley 2004

Choi SR, Howell OW, Carassiti D, Magliozzi R, Gveric D, Muraro PA, et al. Meningeal inflammation plays a role in the pathology of primary progressive multiple sclerosis. *Brain* 2012; 135: 2925–2937.

Confavreux C, Vukusic S. The clinical course of multiple sclerosis. *Handb. Clin. Neurol.* 2014; 122: 343–369.

Curran PJ, Obeidat K, Losardo D. Twelve Frequently Asked Questions About Growth Curve Modeling. *J. Cogn. Dev.* 2010; 11: 121–136.

Das SR, Avants BB, Grossman M, Gee JC. Registration based cortical thickness measurement. *Neuroimage* 2009; 45: 867–79.

Daunizeau J, David O, Stephan KE. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *NeuroImage* 2011; 58: 312–322.

De Stefano N, Giorgio A, Battaglini M, Rovaris M, Sormani MP, Barkhof F, et al. Assessing brain atrophy rates in a large population of untreated multiple sclerosis subtypes. *Neurology* 2010; 74: 1868–1876.

De Stefano N, Stromillo ML, Giorgio A, Bartolozzi ML, Battaglini M, Baldini M, et al. Establishing pathological cut-offs of brain atrophy rates in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 2015: jnnp-2014-309903.

DeLuca GC, Williams K, Evangelou N, Ebers GC, Esiri MM. The contribution of demyelination to axonal loss in multiple sclerosis. *Brain J. Neurol.* 2006; 129: 1507–1516.

Devanand DP, Bansal R, Liu J, Hao X, Pradhaban G, Peterson BS. MRI hippocampal and entorhinal cortex mapping in predicting conversion to Alzheimer's disease. *NeuroImage* 2012; 60: 1622–1629.

Douaud G, Refsum H, de Jager CA, Jacoby R, E. Nichols T, Smith SM, et al. Preventing Alzheimer's disease-related gray matter atrophy by B-vitamin treatment. *Proc. Natl. Acad. Sci.* 2013; 110: 9523–9528.

Dr. T. Jock Murray M. Multiple Sclerosis: The History of a Disease [Internet]. Springer Publishing Company; 2004. Available from: https://books.google.co.uk/books?id=b_g9Nr51T7YC

Duan Y, Liu Y, Liang P, Jia X, Yu C, Qin W, et al. Comparison of grey matter atrophy between patients with neuromyelitis optica and multiple sclerosis: a voxel-based morphometry study. *Eur J Radiol* 2012; 81: e110-4.

- Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. *Neurology* 2000; 55: 1621–1626.
- Eloyan A, Muschelli J, Nebel MB, Liu H, Han F, Zhao T, et al. Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. *Front. Syst. Neurosci.* 2012; 6: 61.
- Eshaghi A, Bodini B, Ridgway GR, García-Lorenzo D, Tozer DJ, Sahraian MA, et al. Temporal and spatial evolution of grey matter atrophy in primary progressive multiple sclerosis. *NeuroImage* 2014; 86: 257–264.
- Eshaghi A, Riyahi-Alam S, Saeedi R, Roostaei T, Nazeri A, Aghsaei A, et al. Classification algorithms with multi-modal data fusion could accurately distinguish neuromyelitis optica from multiple sclerosis. *Neuroimage Clin* 2015; 7: 306–14.
- Filippi M, Rocca MA, Ciccarelli O, De Stefano N, Evangelou N, Kappos L, et al. MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines. *Lancet Neurol.* 2016; 15: 292–303.
- Filippi M, Rovaris M, Inglese M, Barkhof F, De Stefano N, Smith S, et al. Interferon beta-1a for brain tissue loss in patients at presentation with syndromes suggestive of multiple sclerosis: a randomised, double-blind, placebo-controlled trial. *Lancet Lond. Engl.* 2004; 364: 1489–1496.
- Finke C, Heine J, Pache F, Lacheta A, Borisow N, Kuchling J, et al. Normal volumes and microstructural integrity of deep gray matter structures in AQP4+ NMOSD. *Neurol Neuroimmunol Neuroinflamm* 2016; 3: e229.
- Fisher E, Lee J-C, Nakamura K, Rudick RA. Gray matter atrophy in multiple sclerosis: a longitudinal study. *Ann. Neurol.* 2008; 64: 255–265.
- Fisniku LK, Brex PA, Altmann DR, Miszkil KA, Benton CE, Lanyon R, et al. Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain J. Neurol.* 2008; 131: 808–817.
- Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis* [Internet]. Wiley; 2011. Available from: <https://books.google.co.uk/books?id=qOmxRtdNJpEC>
- Fonteijn HM, Modat M, Clarkson MJ, Barnes J, Lehmann M, Hobbs NZ, et al. An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease. *NeuroImage* 2012; 60: 1880–1889.
- Fox NC, Ridgway GR, Schott JM. Algorithms, atrophy and Alzheimer's disease: Cautionary tales for clinical trials. *NeuroImage* 2011; 57: 15–18.
- Franke K, Hagemann G, Schleussner E, Gaser C. Changes of individual BrainAGE during the course of the menstrual cycle. *NeuroImage* 2015; 115: 1–6.
- Fraser MA, Shaw ME, Cherbuin N. A systematic review and meta-analysis of longitudinal hippocampal atrophy in healthy human ageing. *NeuroImage* 2015; 112: 364–374.
- Freeborough PA, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans. Med. Imaging* 1997; 16: 623–629.

Freeman L, Garcia-Lorenzo D, Bottin L, Leroy C, Louapre C, Bodini B, et al. The neuronal component of gray matter damage in multiple sclerosis: A [^{11}C]flumazenil positron emission tomography study: [^{11}C]Flumazenil PET and Neuronal Damage in MS. *Ann. Neurol.* 2015; 78: 554–567.

Frischer JM, Bramow S, Dal-Bianco A, Lucchinetti CF, Rauschka H, Schmidbauer M, et al. The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain J. Neurol.* 2009; 132: 1175–1189.

Friston KJ, Harrison L, Penny W. Dynamic causal modelling. *NeuroImage* 2003; 19: 1273–1302.

Friston KJ, Preller KH, Mathys C, Cagnan H, Heinzle J, Razi A, et al. Dynamic causal modelling revisited. *NeuroImage* 2017

Frost C, Kenward MG, Fox NC. The analysis of repeated ‘direct’ measures of change illustrated with an application in longitudinal imaging. *Stat. Med.* 2004; 23: 3275–3286.

Geraldes R, Esiri MM, DeLuca GC, Palace J. Age-related small vessel disease: a potential contributor to neurodegeneration in multiple sclerosis. *Brain Pathol.* 2017; 27: 707–722.

Geurts JJG, Calabrese M, Fisher E, Rudick RA. Measurement and clinical effect of grey matter pathology in multiple sclerosis. *Lancet Neurol.* 2012; 11: 1082–1092.

Gilmore CP, Donaldson I, Bö L, Owens T, Lowe J, Evangelou N. Regional variations in the extent and pattern of grey matter demyelination in multiple sclerosis: a comparison between the cerebral cortex, cerebellar cortex, deep grey matter nuclei and the spinal cord. *J. Neurol. Neurosurg. Psychiatry* 2009; 80: 182–187.

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks [Internet]. *ArXiv14062661 Cs Stat* 2014 Available from: <http://arxiv.org/abs/1406.2661>

Gronwall DMA. Paced Auditory Serial-Addition Task: A Measure of Recovery from Concussion. *Percept. Mot. Skills* 1977; 44: 367–373.

Groth-Marnat G, Teal M. Block design as a measure of everyday spatial ability: a study of ecological validity. *Percept. Mot. Skills* 2000; 90: 522–526.

Haider L, Simeonidou C, Steinberger G, Hametner S, Grigoriadis N, Deretzi G, et al. Multiple sclerosis deep grey matter: the relation between demyelination, neurodegeneration, inflammation and iron [Internet]. *J Neurol Neurosurg Psychiatry* 2014 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24899728>

Haider L, Zrzavy T, Hametner S, Höftberger R, Bagnato F, Grabner G, et al. The topography of demyelination and neurodegeneration in the multiple sclerosis brain. *Brain J. Neurol.* 2016a; 139: 807–815.

Haider L, Zrzavy T, Hametner S, Höftberger R, Bagnato F, Grabner G, et al. The topography of demyelination and neurodegeneration in the multiple sclerosis brain. *Brain J. Neurol.* 2016b; 139: 807–815.

Hametner S, Wimmer I, Haider L, Pfeifenbring S, Brück W, Lassmann H. Iron and neurodegeneration in the multiple sclerosis brain. *Ann. Neurol.* 2013; 74: 848–861.

- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition. 2nd ed. 2009. Corr. 7th printing 2013 edition. New York, NY: Springer; 2011.
- Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung H-P, Hemmer B, et al. Ocrelizumab versus Interferon Beta-1a in Relapsing Multiple Sclerosis. *N. Engl. J. Med.* 2017; 376: 221–234.
- Hawker K, O'Connor P, Freedman MS, Calabresi PA, Antel J, Simon J, et al. Rituximab in patients with primary progressive multiple sclerosis: results of a randomized double-blind placebo-controlled multicenter trial. *Ann. Neurol.* 2009; 66: 460–471.
- Healy BC, Engler D, Glanz B, Musallam A, Chitnis T. Assessment of definitions of sustained disease progression in relapsing-remitting multiple sclerosis. *Mult. Scler. Int.* 2013; 2013: 189624.
- Hedman AM, van Haren NE, Schnack HG, Kahn RS, Hulshoff Pol HE. Human brain changes across the life span: a review of 56 longitudinal magnetic resonance imaging studies. *Hum Brain Mapp* 2012; 33: 1987–2002.
- Henneman WJP, Sluimer JD, Barnes J, van der Flier WM, Sluimer IC, Fox NC, et al. Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures. *Neurology* 2009; 72: 999–1007.
- Hidalgo B, Goodman M. Multivariate or Multivariable Regression? *Am. J. Public Health* 2013; 103: 39–40.
- Hobart J, Lamping D, Fitzpatrick R, Riazi A, Thompson A. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain J. Neurol.* 2001; 124: 962–973.
- Hobbs NZ, Henley SMD, Wild EJ, Leung KK, Frost C, Barker RA, et al. Automated quantification of caudate atrophy by local registration of serial MRI: evaluation and application in Huntington's disease. *NeuroImage* 2009; 47: 1659–1665.
- Howell OW, Reeves CA, Nicholas R, Carassiti D, Radotra B, Gentleman SM, et al. Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis. *Brain J. Neurol.* 2011; 134: 2755–2771.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Model. Multidiscip. J.* 1999; 6: 1–55.
- Huppertz H-J, Möller L, Südmeyer M, Hilker R, Hattingen E, Egger K, et al. Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification. *Mov. Disord. Off. J. Mov. Disord. Soc.* 2016
- Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. *Am. Polit. Sci. Rev.* 2011; 105: 765–789.
- International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CCA, et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* 2011; 476: 214–219.

- Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 2001; 5: 143–156.
- Jones BC, Nair G, Shea CD, Crainiceanu CM, Cortese ICM, Reich DS. Quantification of multiple-sclerosis-related brain atrophy in two heterogeneous MRI datasets using mixed-effects modeling. *NeuroImage Clin.* 2013; 3: 171–179.
- Jorge Cardoso M, Leung K, Modat M, Keihaninejad S, Cash D, Barnes J, et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Med. Image Anal.* 2013; 17: 671–684.
- Josephs KA, Dickson DW, Tosakulwong N, Weigand SD, Murray ME, Petrucelli L, et al. Rates of hippocampal atrophy and presence of post-mortem TDP-43 in patients with Alzheimer's disease: a longitudinal retrospective study. *Lancet Neurol.* 2017; 16: 917–924.
- Jürgens T, Jafari M, Kreutzfeldt M, Bahn E, Brück W, Kerschensteiner M, et al. Reconstruction of single cortical projection neurons reveals primary spine loss in multiple sclerosis. *Brain J. Neurol.* 2016; 139: 39–46.
- Kalincik T, Brown JWL, Robertson N, Willis M, Scolding N, Rice CM, et al. Treatment effectiveness of alemtuzumab compared with natalizumab, fingolimod, and interferon beta in relapsing-remitting multiple sclerosis: a cohort study. *Lancet Neurol.* 2017; 16: 271–281.
- Kalkers NF, Ameziane N, Bot JCJ, Minneboo A, Polman CH, Barkhof F. Longitudinal brain volume measurement in multiple sclerosis: rate of brain atrophy is independent of the disease subtype. *Arch. Neurol.* 2002; 59: 1572–1576.
- Kappos L, Bar-Or A, Cree BAC, Fox RJ, Giovannoni G, Gold R, et al. Siponimod versus placebo in secondary progressive multiple sclerosis (EXPAND): a double-blind, randomised, phase 3 study. *Lancet Lond. Engl.* 2018; 391: 1263–1273.
- Kappos L, De Stefano N, Freedman MS, Cree BA, Radue E-W, Sprenger T, et al. Inclusion of brain volume loss in a revised measure of 'no evidence of disease activity' (NEDA-4) in relapsing-remitting multiple sclerosis. *Mult. Scler. Houndmills Basingstoke Engl.* 2015
- Kappos L, Edan G, Freedman MS, Montalbán X, Hartung H-P, Hemmer B, et al. The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. *Neurology* 2016
- Keshavan A, Paul F, Beyer MK, Zhu AH, Papinutto N, Shinohara RT, et al. Power estimation for non-standardized multisite studies. *NeuroImage* 2016; 134: 281–294.
- Kievit RA, Davis SW, Mitchell DJ, Taylor JR, Duncan J, Team C-CR, et al. Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nat. Commun.* 2014; 5: 5658.
- Kilsdonk ID, Jonkman LE, Klaver R, van Veluw SJ, Zwanenburg JJM, Kuijer JPA, et al. Increased cortical grey matter lesion detection in multiple sclerosis with 7 T MRI: a post-mortem verification study. *Brain* 2016; 139: 1472–1481.
- Kim HJ, Paul F, Lana-Peixoto MA, Tenenbaum S, Asgari N, Palace J, et al. MRI characteristics of neuromyelitis optica spectrum disorder: an international update. *Neurology* 2015; 84: 1165–73.

- Kipps CM, Duggins AJ, Mahant N, Gomes L, Ashburner J, McCusker EA. Progression of structural neuropathology in preclinical Huntington's disease: a tensor based morphometry study. *J. Neurol. Neurosurg. Psychiatry* 2005; 76: 650–655.
- Kister I, Herbert J, Zhou Y, Ge Y. Ultrahigh-Field MR (7 T) Imaging of Brain Lesions in Neuromyelitis Optica. *Mult Scler Int* 2013; 2013: 398259.
- Klein A, Tourville J. 101 labeled brain images and a consistent human cortical labeling protocol [Internet]. *Front. Neurosci.* 2012; 6[cited 2016 Jul 7] Available from: <http://journal.frontiersin.org/article/10.3389/fnins.2012.00171/abstract>
- Kline RB. Principles and Practice of Structural Equation Modeling, Fourth Edition [Internet]. Guilford Publications; 2015. Available from: <https://books.google.co.uk/books?id=3VauCgAAQBAJ>
- Klöppel S, Abdulkadir A, Jack CR, Koutsouleris N, Mourão-Miranda J, Vemuri P. Diagnostic neuroimaging across diseases. *NeuroImage* 2012; 61: 457–463.
- Klöppel S, Chu C, Tan GC, Draganski B, Johnson H, Paulsen JS, et al. Automatic detection of preclinical neurodegeneration: presymptomatic Huntington disease. *Neurology* 2009; 72: 426–431.
- Klöppel Stefan, Stonnington CM, Barnes J, Chen F, Chu C, Good CD, et al. Accuracy of dementia diagnosis: a direct comparison between radiologists and a computerized method. *Brain J. Neurol.* 2008; 131: 2969–2974.
- Klöppel S., Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, et al. Automatic classification of MR scans in Alzheimer's disease. *Brain* 2008; 131: 681–9.
- Kolasinski J, Stagg CJ, Chance SA, Deluca GC, Esiri MM, Chang E-H, et al. A combined post-mortem magnetic resonance imaging and quantitative histological study of multiple sclerosis pathology. *Brain J. Neurol.* 2012; 135: 2938–2951.
- Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983; 33: 1444–1452.
- Kutzelnigg A, Faber-Rod JC, Bauer J, Lucchinetti CF, Sorensen PS, Laursen H, et al. Widespread demyelination in the cerebellar cortex in multiple sclerosis. *Brain Pathol. Zurich Switz.* 2007; 17: 38–44.
- Kutzelnigg A, Lucchinetti CF, Stadelmann C, Brück W, Rauschka H, Bergmann M, et al. Cortical demyelination and diffuse white matter injury in multiple sclerosis. *Brain J. Neurol.* 2005; 128: 2705–2712.
- Lansley J, Mataix-Cols D, Grau M, Radua J, Sastre-Garriga J. Localized grey matter atrophy in multiple sclerosis: a meta-analysis of voxel-based morphometry studies and associations with functional disability. *Neurosci. Biobehav. Rev.* 2013; 37: 819–830.
- Larochelle C, Uphaus T, Prat A, Zipp F. Secondary Progression in Multiple Sclerosis: Neuronal Exhaustion or Distinct Pathology? *Trends Neurosci.* 2016; 39: 325–339.
- Lebedev AV, Westman E, Van Westen GJP, Kramberger MG, Lundervold A, Aarsland D, et al. Random Forest ensembles for detection and prediction of Alzheimer's disease with a good between-cohort robustness. *NeuroImage Clin.* 2014; 6: 115–125.

Lennon VA, Wingerchuk DM, Kryzer TJ, Pittock SJ, Lucchinetti CF, Fujihara K, et al. A serum autoantibody marker of neuromyelitis optica: distinction from multiple sclerosis. *Lancet Lond. Engl.* 2004; 364: 2106–2112.

Leung KK, Ridgway GR, Ourselin S, Fox NC, Alzheimer's Disease Neuroimaging I. Consistent multi-time-point brain atrophy estimation from the boundary shift integral. *Neuroimage* 2012; 59: 3995–4005.

Lindberg RLP, De Groot CJA, Certa U, Ravid R, Hoffmann F, Kappos L, et al. Multiple sclerosis as a generalized CNS disease--comparative microarray analysis of normal appearing white matter and lesions in secondary progressive MS. *J. Neuroimmunol.* 2004; 152: 154–167.

Liu Y, Fu Y, Schoonheim MM, Zhang N, Fan M, Su L, et al. Structural MRI substrates of cognitive impairment in neuromyelitis optica. *Neurology* 2015; 85: 1491–9.

Liu Y, Wang J, Daams M, Weiler F, Hahn HK, Duan Y, et al. Differential patterns of spinal cord and brain atrophy in NMO and MS. *Neurology* 2015; 84: 1465–72.

Lopez-Martinez D, Picard R. Wearable Technologies for Multiple Sclerosis: The Future Role of Wearable Stress Measurement in Improving Quality of Life. In: *Second International Conference on Smart Portable, Wearable, Implantable and Disability-Oriented Devices and Systems (SPWID'16)*. Valencia, Spain: 2016.

Lorscheider J, Buzzard K, Jokubaitis V, Spelman T, Havrdova E, Horakova D, et al. Defining secondary progressive multiple sclerosis. *Brain J. Neurol.* 2016; 139: 2395–2405.

Lorscheider J, Jokubaitis VG, Spelman T, Izquierdo G, Lugaresi A, Havrdova E, et al. Anti-inflammatory disease-modifying treatment and short-term disability progression in SPMS. *Neurology* 2017a; 89: 1050–1059.

Lorscheider J, Jokubaitis VG, Spelman T, Izquierdo G, Lugaresi A, Havrdova E, et al. Anti-inflammatory disease-modifying treatment and short-term disability progression in SPMS. *Neurology* 2017b; 89: 1050–1059.

Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology* 1996; 46: 907–911.

Lucchinetti CF, Guo Y, Popescu BF, Fujihara K, Itoyama Y, Misu T. The pathology of an autoimmune astrocytopathy: lessons learned from neuromyelitis optica. *Brain Pathol* 2014; 24: 83–97.

Lucchinetti CF, Popescu BFG, Bunyan RF, Moll NM, Roemer SF, Lassmann H, et al. Inflammatory Cortical Demyelination in Early Multiple Sclerosis. *N. Engl. J. Med.* 2011; 365: 2188–2197.

Lui YW, Xue Y, Kenul D, Ge Y, Grossman RI, Wang Y. Classification algorithms using multiple MRI features in mild traumatic brain injury. *Neurology* 2014; 83: 1235–40.

Lukas C, Minneboo A, de Groot V, Moraal B, Knol DL, Polman CH, et al. Early central atrophy rate predicts 5 year clinical outcome in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 2010; 81: 1351–1356.

- Magliozzi R, Howell O, Vora A, Serafini B, Nicholas R, Puopolo M, et al. Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology. *Brain* 2006; 130: 1089–1104.
- Magliozzi R, Howell OW, Reeves C, Roncaroli F, Nicholas R, Serafini B, et al. A Gradient of neuronal loss and meningeal inflammation in multiple sclerosis. *Ann. Neurol.* 2010; 68: 477–493.
- Mahad DH, Trapp BD, Lassmann H. Pathological mechanisms in progressive multiple sclerosis. *Lancet Neurol.* 2015; 14: 183–193.
- Mainero C, Louapre C, Govindarajan ST, Giannini C, Nielsen AS, Cohen-Adad J, et al. A gradient in cortical pathology in multiple sclerosis by in vivo quantitative 7 T imaging. *Brain* 2015; 138: 932–945.
- Mallik S, Muhlert N, Samson RS, Sethi V, Wheeler-Kingshott CAM, Miller DH, et al. Regional patterns of grey matter atrophy and magnetisation transfer ratio abnormalities in multiple sclerosis clinical subgroups: a voxel-based analysis study. *Mult. Scler. Houndmills Basingstoke Engl.* 2015; 21: 423–432.
- Malone IB, Leung KK, Clegg S, Barnes J, Whitwell JL, Ashburner J, et al. Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance. *NeuroImage* 2015; 104: 366–372.
- Marrie RA. Comorbidity in multiple sclerosis: implications for patient care. *Nat. Rev. Neurol.* 2017; 13: 375–382.
- Marrie RA, Horwitz RJ. Emerging effects of comorbidities on multiple sclerosis. *Lancet Neurol.* 2010; 9: 820–828.
- Matthews L, Kolind S, Brazier A, Leite MI, Brooks J, Traboulsee A, et al. Imaging Surrogates of Disease Activity in Neuromyelitis Optica Allow Distinction from Multiple Sclerosis. *PLoS One* 2015; 10: e0137715.
- Matthews PM. Decade in review-multiple sclerosis: new drugs and personalized medicine for multiple sclerosis. *Nat. Rev. Neurol.* 2015; 11: 614–616.
- McDonald WI, Compston A, Edan G, Goodkin D, Hartung H-P, Lublin FD, et al. Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Ann. Neurol.* 2001; 50: 121–127.
- Melão A. MSBase, Icometrix Collaborate on Global Imaging Project to Understand MS Progression [Internet]. *Mult. Scler. News Today* 2017[cited 2018 Jan 1] Available from: <https://multiplesclerosisnewstoday.com/2017/07/05/msbase-icometrix/>
- Miller DH, Chard DT, Ciccarelli O. Clinically isolated syndromes. *Lancet Neurol.* 2012; 11: 157–169.
- Minagar A, Barnett MH, Benedict RH, Pelletier D, Pirko I, Sahraian MA, et al. The thalamus and multiple sclerosis: modern views on pathologic, imaging, and clinical aspects. *Neurology* 2013; 80: 210–9.
- Montalban X, Hauser SL, Kappos L, Arnold DL, Bar-Or A, Comi G, et al. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *N. Engl. J. Med.* 2017; 376: 209–220.

- Moon Y, McGinnis RS, Seagers K, Motl RW, Sheth N, Wright JA, et al. Monitoring gait in multiple sclerosis with novel wearable motion sensors. *PLOS ONE* 2017; 12: e0171346.
- Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Alzheimer's Disease Neuroimaging Initiative. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. *NeuroImage* 2015; 104: 398–412.
- van Munster CEP, Uitdehaag BMJ. Outcome Measures in Clinical Trials for Multiple Sclerosis. *CNS Drugs* 2017; 31: 217–236.
- Nakamura K, Brown RA, Narayanan S, Collins DL, Arnold DL, Alzheimer's Disease Neuroimaging Initiative. Diurnal fluctuations in brain volume: Statistical analyses of MRI from large populations. *NeuroImage* 2015; 118: 126–132.
- Nie D, Trullo R, Petitjean C, Ruan S, Shen D. Medical Image Synthesis with Context-Aware Generative Adversarial Networks [Internet]. 2016[cited 2017 Dec 31] Available from: <https://arxiv.org/abs/1612.05362>
- Oliphant TE. Python for Scientific Computing. *Comput. Sci. Eng.* 2007; 9: 10–20.
- Ontaneda D, Thompson AJ, Fox RJ, Cohen JA. Progressive multiple sclerosis: prospects for disease therapy, repair, and restoration of function. *The Lancet* 2017; 389: 1357–1366.
- Optic Neuritis Study Group. Multiple sclerosis risk after optic neuritis: final optic neuritis treatment trial follow-up. *Arch. Neurol.* 2008; 65: 727–732.
- Oxtoby NP, Alexander DC, EuroPOND consortium. Imaging plus X: multimodal models of neurodegenerative disease. *Curr. Opin. Neurol.* 2017; 30: 371–379.
- Oxtoby NP, Garbarino S, Firth NC, Warren JD, Schott JM, Alexander DC, et al. Data-Driven Sequence of Changes to Anatomical Brain Connectivity in Sporadic Alzheimer's Disease. *Front. Neurol.* 2017; 8: 580.
- Oxtoby NP, Young AL, Cash DM, Benzinger TLS, Fagan AM, Morris JC, et al. Data-driven models of dominantly-inherited Alzheimer's disease progression. *Brain J. Neurol.* 2018; 141: 1529–1544.
- Paling D, Solanky BS, Riemer F, Tozer DJ, Wheeler-Kingshott CAM, Kapoor R, et al. Sodium accumulation is associated with disability and a progressive course in multiple sclerosis. *Brain J. Neurol.* 2013; 136: 2305–2317.
- Pardini M, Sudre CH, Prados F, Yaldizli Ö, Sethi V, Muhlert N, et al. Relationship of grey and white matter abnormalities with distance from the surface of the brain in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 2016; 87: 1212–1217.
- Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* 2011; 56: 907–22.
- Paviour DC, Price SL, Jahanshahi M, Lees AJ, Fox NC. Longitudinal MRI in progressive supranuclear palsy and multiple system atrophy: rates and regions of atrophy. *Brain* 2006; 129: 1040–1049.
- Peelle JE, Cusack R, Henson RN. Adjusting for global effects in voxel-based morphometry: gray matter decline in normal aging. *Neuroimage* 2012; 60: 1503–16.

Penny WD, Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE. Statistical Parametric Mapping: The Analysis of Functional Brain Images [Internet]. Elsevier Science; 2011. Available from: https://books.google.co.uk/books?id=G_qdEsDlkp0C

Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage* 2009; 45: S199–S209.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models [Internet]. 2017[cited 2017 Oct 18] Available from: <https://cran.r-project.org/web/packages/nlme/citation.html>

Pinheiro JC, Bates D. Mixed-Effects Models in S and S-PLUS [Internet]. Springer; 2009. Available from: <https://books.google.co.uk/books?id=y54QDUTmvDcC>

Pittock SJ, Weinshenker BG, Lucchinetti CF, Wingerchuk DM, Corboy JR, Lennon VA. Neuromyelitis optica brain lesions localized at sites of high aquaporin 4 expression. *Arch Neurol* 2006; 63: 964–8.

Polman CH, Reingold SC, Banwell B, Clanet M, Cohen JA, Filippi M, et al. Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Ann. Neurol.* 2011; 69: 292–302.

Polman CH, Reingold SC, Edan G, Filippi M, Hartung HP, Kappos L, et al. Diagnostic criteria for multiple sclerosis: 2005 revisions to the ‘McDonald Criteria’. *Ann Neurol* 2005; 58: 840–6.

Popescu V, Agosta F, Hulst HE, Sluimer IC, Knol DL, Sormani MP, et al. Brain atrophy and lesion load predict long term disability in multiple sclerosis. *J. Neurol. Neurosurg. Psychiatry* 2013; 84: 1082–1091.

Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann. Neurol.* 1983; 13: 227–231.

Prados F, Cardoso MJ, Kanber B, Ciccarelli O, Kapoor R, Gandini Wheeler-Kingshott CAM, et al. A multi-time-point modality-agnostic patch-based method for lesion filling in multiple sclerosis. *NeuroImage* 2016; 139: 376–384.

Prados F, Cardoso MJ, Leung KK, Cash DM, Modat M, Fox NC, et al. Measuring brain atrophy with a generalized formulation of the boundary shift integral. *Neurobiol. Aging* 2015; 36 Suppl 1: S81-90.

Preziosa P, Pagani E, Mesaros S, Riccitelli GC, Dackovic J, Drulovic J, et al. Progression of regional atrophy in the left hemisphere contributes to clinical and cognitive deterioration in multiple sclerosis: A 5-year study. *Hum. Brain Mapp.* 2017

R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available from: <http://www.R-project.org/>

Raichle ME. The brain’s default mode network. *Annu. Rev. Neurosci.* 2015; 38: 433–447.

Ransohoff RM. How neuroinflammation contributes to neurodegeneration. *Science* 2016; 353: 777–783.

- Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Bürger C, et al. Prediction of Individual Response to Electroconvulsive Therapy via Machine Learning on Structural Magnetic Resonance Imaging Data. *JAMA Psychiatry* 2016; 73: 557–564.
- Reuter M, Fischl B. Avoiding asymmetry-induced bias in longitudinal image processing. *NeuroImage* 2011; 57: 19–21.
- Reuter M, Rosas HD, Fischl B. Highly accurate inverse consistent registration: a robust approach. *NeuroImage* 2010; 53: 1181–1196.
- Reuter M, Schmansky NJ, Rosas HD, Fischl B. Within-subject template estimation for unbiased longitudinal image analysis. *NeuroImage* 2012; 61: 1402–1418.
- Reuter M, Tisdall MD, Qureshi A, Buckner RL, van der Kouwe AJW, Fischl B. Head motion during MRI acquisition reduces gray matter volume and thickness estimates. *NeuroImage* 2015; 107: 107–115.
- Riccitelli G, Rocca MA, Pagani E, Rodegher ME, Rossi P, Falini A, et al. Cognitive impairment in multiple sclerosis is associated to different patterns of gray matter atrophy according to clinical phenotype. *Hum. Brain Mapp.* 2011; 32: 1535–1543.
- Rocca MA, Battaglini M, Benedict RHB, De Stefano N, Geurts JJG, Henry RG, et al. Brain MRI atrophy quantification in MS: From methods to clinical application. *Neurology* 2017; 88: 403–413.
- Rocca MA, Mesaros S, Pagani E, Sormani MP, Comi G, Filippi M. Thalamic damage and long-term progression of disability in multiple sclerosis. *Radiology* 2010; 257: 463–469.
- Roosendaal SD, Bendfeldt K, Vrenken H, Polman CH, Borgwardt S, Radue EW, et al. Grey matter volume in a large cohort of MS patients: relation to MRI parameters and disability. *Mult. Scler. J.* 2011; 17: 1098–1106.
- Rosseel Y. **lavaan** : An R Package for Structural Equation Modeling [Internet]. *J. Stat. Softw.* 2012; 48[cited 2017 Aug 26] Available from: <http://www.jstatsoft.org/v48/i02/>
- Ruggieri S, Petracca M, Miller A, Krieger S, Ghassemi R, Bencosme Y, et al. Association of deep gray matter damage with cortical and spinal cord degeneration in primary progressive multiple sclerosis. *JAMA Neurol.* 2015; 72: 1466–1474.
- Sabuncu MR, Bernal-Rusiel JL, Reuter M, Greve DN, Fischl B. Event time analysis of longitudinal neuroimage data. *NeuroImage* 2014; 97: 9–18.
- Sailer M, Fischl B, Salat D, Tempelmann C, Schonfeld MA, Busa E, et al. Focal thinning of the cerebral cortex in multiple sclerosis. *Brain* 2003; 126: 1734–44.
- Saji E, Arakawa M, Yanagawa K, Toyoshima Y, Yokoseki A, Okamoto K, et al. Cognitive impairment and cortical degeneration in neuromyelitis optica. *Ann. Neurol.* 2012; n/a-n/a.
- Sato DK, Callegaro D, Lana-Peixoto MA, Waters PJ, de Haidar Jorge FM, Takahashi T, et al. Distinction between MOG antibody-positive and AQP4 antibody-positive NMO spectrum disorders. *Neurology* 2014; 82: 474–81.
- Schmidt P, Gaser C, Arsic M, Buck D, Förchler A, Berthele A, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. *NeuroImage* 2012; 59: 3774–3783.

Schoonheim MM, Hulst HE, Brandt RB, Strik M, Wink AM, Uitdehaag BMJ, et al. Thalamus structure and function determine severity of cognitive impairment in multiple sclerosis. *Neurology* 2015; 84: 776–783.

Schumacher GA, Beebe G, Kibler RF, Kurland LT, Kurtzke JF, McDowell F, et al. Problems of experimental trials of therapy in multiple sclerosis: report by the panel on the evaluation of experimental trials of therapy in multiple sclerosis. *Ann. N. Y. Acad. Sci.* 1965; 122: 552–568.

Sdika M, Pelletier D. Nonrigid registration of multiple sclerosis brain images using lesion inpainting for morphometry or lesion mapping. *Hum Brain Mapp* 2009; 30: 1060–7.

Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. *Neuron* 2009; 62: 42–52.

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, et al. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage* 2008; 39: 1064–80.

Shimizu J, Hatanaka Y, Hasegawa M, Iwata A, Sugimoto I, Date H, et al. IFN β -1b may severely exacerbate Japanese optic-spinal MS in neuromyelitis optica spectrum. *Neurology* 2010; 75: 1423–1427.

Singer JD, Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence* [Internet]. Oxford University Press; 2003. Available from: <https://books.google.co.uk/books?id=mGwUjlvNiSoC>

Sinnecker T, Dorr J, Pfueller CF, Harms L, Ruprecht K, Jarius S, et al. Distinct lesion morphology at 7-T MRI differentiates neuromyelitis optica from multiple sclerosis. *Neurology* 2012; 79: 708–14.

Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002; 17: 143–55.

Smith SM, De Stefano N, Jenkinson M, Matthews PM. Normalized accurate measurement of longitudinal brain change. *J. Comput. Assist. Tomogr.* 2001; 25: 466–475.

Smith SM, Rao A, De Stefano N, Jenkinson M, Schott JM, Matthews PM, et al. Longitudinal and cross-sectional analysis of atrophy in Alzheimer's disease: Cross-validation of BSI, SIENA and SIENAX. *NeuroImage* 2007; 36: 1200–1206.

Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, et al. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *NeuroImage* 2002; 17: 479–489.

Sormani MP, Arnold DL, De Stefano N. Treatment effect on brain atrophy correlates with treatment effect on disability in multiple sclerosis. *Ann. Neurol.* 2014; 75: 43–49.

Steenwijk MD, Geurts JJG, Daams M, Tijms BM, Wink AM, Balk LJ, et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain J. Neurol.* 2016; 139: 115–126.

Steenwijk MD, Pouwels PJW, Daams M, van Dalen JW, Caan MWA, Richard E, et al. Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (kNN-TTPs). *NeuroImage Clin.* 2013; 3: 462–469.

- Stephan KE, Mathys C. Computational approaches to psychiatry. *Curr. Opin. Neurobiol.* 2014; 25: 85–92.
- Stys PK, Zamponi GW, van Minnen J, Geurts JJG. Will the real multiple sclerosis please stand up? *Nat. Rev. Neurosci.* 2012; 13: 507–514.
- Sudre CH, Cardoso MJ, Bouvy WH, Biessels GJ, Barnes J, Ourselin S. Bayesian model selection for pathological neuroimaging data applied to white matter lesion segmentation. *IEEE Trans. Med. Imaging* 2015; 34: 2079–2102.
- Sudre CH, Cardoso MJ, Ourselin S, Alzheimer's Disease Neuroimaging Initiative. Longitudinal segmentation of age-related white matter hyperintensities. *Med. Image Anal.* 2017; 38: 50–64.
- Tabrizi SJ, Reilmann R, Roos RA, Durr A, Leavitt B, Owen G, et al. Potential endpoints for clinical trials in premanifest and early Huntington's disease in the TRACK-HD study: analysis of 24 month observational data. *Lancet Neurol.* 2012; 11: 42–53.
- Takao H, Hayashi N, Ohtomo K. Effect of scanner in longitudinal studies of brain volume changes. *J. Magn. Reson. Imaging JMRI* 2011; 34: 438–444.
- Thompson AJ. A much-needed focus on progression in multiple sclerosis. *Lancet Neurol.* 2015; 14: 133–135.
- Thompson AJ. Challenge of progressive multiple sclerosis therapy. *Curr. Opin. Neurol.* 2017
- Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol.* 2017
- Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* 1996; 58: 267–288.
- Trapp BD, Stys PK. Virtual hypoxia and chronic necrosis of demyelinated axons in multiple sclerosis. *Lancet Neurol.* 2009; 8: 280–291.
- Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 2010; 29: 1310–1320.
- Tustison NJ, Cook PA, Klein A, Song G, Das SR, Duda JT, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage* 2014; 99: 166–79.
- University of California, San Francisco MS-EPIC Team:, Cree BAC, Gourraud P-A, Oksenberg JR, Bevan C, Crabtree-Hartman E, et al. Long-term evolution of multiple sclerosis disability in the treatment era. *Ann. Neurol.* 2016
- Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models* [Internet]. Springer New York; 2006. Available from: <https://books.google.co.uk/books?id=tGw-9HRV2UEC>
- Vrenken H, Vos EK, van der Flier WM, Sluimer IC, Cover KS, Knol DL, et al. Validation of the automated method VIENA: an accurate, precise, and robust measure of ventricular enlargement. *Hum. Brain Mapp.* 2014; 35: 1101–1110.
- Wagenmakers E-J, Farrell S. AIC model selection using Akaike weights. *Psychon. Bull. Rev.* 2004; 11: 192–196.

- Wang C, Beadnall HN, Hatton SN, Bader G, Tomic D, Silva DG, et al. Automated brain volumetrics in multiple sclerosis: a step closer to clinical application. *J. Neurol. Neurosurg. Psychiatry* 2016; 87: 754–757.
- Wang Y, Fan Y, Bhatt P, Davatzikos C. High-dimensional pattern regression using machine learning: from medical images to continuous clinical variables. *NeuroImage* 2010; 50: 1519–1535.
- Wechsler D. Wechsler Abbreviated Scale of Intelligence. 2011
- Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, et al. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimers Dement. J. Alzheimers Assoc.* 2017; 13: e1–e85.
- Wijeratne PA, Young AL, Oxtoby NP, Marinescu RV, Firth NC, Johnson EB, et al. An image-based model of brain volume biomarker changes in Huntington's disease. *Ann. Clin. Transl. Neurol.* 2018; 5: 570–582.
- Wijnands JMA, Kingwell E, Zhu F, Zhao Y, Högg T, Stadnyk K, et al. Health-care use before a first demyelinating event suggestive of a multiple sclerosis prodrome: a matched cohort study. *Lancet Neurol.* 2017; 16: 445–451.
- Wingerchuk DM. Diagnosis and treatment of neuromyelitis optica. *The Neurologist* 2007; 13: 2–11.
- Wingerchuk DM, Banwell B, Bennett JL, Cabre P, Carroll W, Chitnis T, et al. International consensus diagnostic criteria for neuromyelitis optica spectrum disorders. *Neurology* 2015; 85: 177–189.
- Wingerchuk DM, Lennon VA, Pittock SJ, Lucchinetti CF, Weinshenker BG. Revised diagnostic criteria for neuromyelitis optica. *Neurology* 2006; 66: 1485–1489.
- Wingerchuk DM, Pittock SJ, Lucchinetti CF, Lennon VA, Weinshenker BG. A secondary progressive clinical course is uncommon in neuromyelitis optica. *Neurology* 2007; 68: 603–605.
- Wingerchuk DM, Weinshenker BG. Neuromyelitis optica (Devic's syndrome). *Handb Clin Neurol* 2014; 122: 581–99.
- Wottschel V, Alexander DC, Kwok PP, Chard DT, Stromillo ML, De Stefano N, et al. Predicting outcome in clinically isolated syndrome using machine learning. *NeuroImage Clin.* 2015; 7: 281–287.
- Young AL, Oxtoby NP, Daga P, Cash DM, Fox NC, Ourselin S, et al. A data-driven model of biomarker changes in sporadic Alzheimer's disease. *Brain J. Neurol.* 2014; 137: 2564–2577.
- Zhang D, Raichle ME. Disease and the brain's dark energy. *Nat. Rev. Neurol.* 2010; 6: 15–28.
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 2001; 20: 45–57.

Zivadinov R, Heininen-Brown M, Schirda CV, Poloni GU, Bergsland N, Magnano CR, et al. Abnormal subcortical deep-gray matter susceptibility-weighted imaging filtered phase measurements in patients with multiple sclerosis: a case-control study. *NeuroImage* 2012; 59: 331–339.