Imaging Markers of Brain Network Disruption in Multiple Sclerosis

Thomas Welton

Thesis submitted to the University of Nottingham

for the Degree of Doctor of Philosophy

April 2017

Abstract

Cognitive impairment and fatigue are prevalent and impactful symptoms of multiple sclerosis (MS). Effective markers are required by clinical studies to accurately test the efficacy of treatments for these symptoms. Graph analysis of brain networks based on magnetic resonance imaging (MRI) data can feasibly provide useful candidate markers of cognitive impairment and fatigability in MS which may be more objective, reliable and specific than existing markers. My original contribution to knowledge is therefore an exposition of the following hypothesis: “summary graph-theoretic descriptors of brain network organisation are good candidate markers of cognition or fatigue in MS”. To achieve this, network metrics were assessed based on three main criteria: reliability (“are the measurements the same across time and settings?”), validity (“do they measure what they are supposed to measure?”) and responsiveness (“are they altered when a change in cognitive state is induced?”).

The applicability of the graph-theoretic approach was first established with a spatial meta-analysis of tract integrity and its relevance to cognition and disability. Reliability over time in healthy subjects was assessed by systematic review and reliability between different scanners and between MS and control groups was assessed in two longitudinal datasets by measuring intra-class correlation (ICC) of graph metrics. The validity criterion was assessed in an analysis of covariance and linear regression of cognitive and fatigue measures with brain network metrics in people with MS. Finally, an

exploration of network dynamics during a sustained attention task with a sliding-window approach was performed to test the immediate responsiveness of the measures to alterations in cognitive state.

Spatial meta-analysis of white matter tract degradation was performed using the Signed Differential Mapping method. Statistical maps were gathered from the original authors of studies which performed voxelwise correlations between fractional anisotropy (a measure of white matter integrity based on diffusion tensor imaging data) and measures of either cognitive impairment or physical disability. The combined sample included 495 people with MS and 253 controls from 12 studies. MS diagnosis was significantly associated with widespread lower tract fractional anisotropy. Distributions of voxels with significantly lower fractional anisotropy in relation to cognition and physical disability were only minimally overlapping. The number of and effect sizes for significant clusters in the cognition comparison were greater than those for the physical disability comparison, suggesting a greater relevance of cerebral white matter damage to cognition. The main results remained significant when using a stringent p-value threshold of 0.00001 to control for false positives.

The next analysis was a systematic review of the reproducibility of graph metrics over time in healthy people. Online databases were searched for articles reporting ICCs for graph metrics based on imaging data and information was recorded on the sample size, acquisition method, inter-scan interval and reported ICCs. Twenty-six articles were included, with a combined sample size of 676. Overall, reproducibility over time was rated as “good”, but heterogeneity of methods precluded in-depth quantitative analysis. A qualitative synthesis of results highlighted the main methodologic factors

affecting reproducibility, which included: ICC type, retest interval, fibre tracking algorithm, graph metric type, image processing strategy, region of interest size, graph threshold and acquisition method.

Reliability of brain network metrics between scanners was tested using a travelling-subjects dataset in which 5 subjects each underwent a resting-state functional MRI scan at 10 sites. Graph metrics were calculated for each scan and then tested for ICC across sites. Reproducibility was “poor” for most metrics (characteristic path length ICC=0.23, global efficiency ICC=0.18, modularity ICC=0.24) and “fair” for two (clustering coefficient ICC=0.43, small-worldness ICC=0.42). There was limited evidence that some subjects tended to produce less reliable results and that magnets with higher field strengths did not produce more reliable results. The main implication is that multi-site studies using graph analysis of brain MRI data should investigate inter-site reproducibility beforehand.

To investigate the validity of graph metrics as markers of cognitive impairment and fatigue, MRI and neurocognitive data were first gathered from 37 people with MS and 23 matched controls. The sample was characterised in detail and comprised a range of cognitive abilities. Data quality was investigated and the small-world structure of the data was confirmed by comparison to random and lattice graphs. Analysis of covariance controlling for age, sex and education showed significant group differences for all but one graph metric. Linear regression models predicted the main measures of cognitive impairment in the MS group, but not in the control group. Measures of fatigue were not well-explained by graph metrics. The direction of the relationships indicated that greater levels of cognitive impairment were related to increased network

clustering and modularity, longer average path lengths, lower small-worldness, lower levels of education, old age and sleep disturbance.

Finally, responsiveness of graph metrics was investigated in an analysis of functional network dynamics during performance of a sustained attention task. A “sliding-window” approach was taken, in which network metrics were calculated for 84 100-second windows at increments along the fMRI timeseries. Reaction times in the task showed a learning effect for both groups, but were consistently slower for the MS group. Plots of graph metrics over time showed differing responses to the task and to the transition between task and rest periods between groups. The small-worldness and clustering coefficient metrics were correlated with reaction times for both MS (small-worldness: r=0.623, <0.001; clustering coefficient: r=0.554, p=<0.001) and control (small-worldness: r=0.586, <0.001; clustering coefficient: r=0.627, p=<0.001) groups, but the characteristic path length metric was not (MS: r=-0.154, p=0.313; control: r=0.343, p=0.021).

Disconnection of cortical areas by degradation of white matter is a viable explanation for cognitive symptoms in MS. There is some evidence that increased network segregation and decreased network integration may explain cognitive symptomatology. Graph theoretic summary brain network metrics do have potential for use as complimentary information to existing markers of cognitive impairment in clinical studies.

List of Publications October 2013 – April 2017

Journal Articles

Meng D, Hosseini AA, Simpson RJ, Welton T, Dineen RA & Auer DP, 2017. On the nature and role of large-scale network dysfunction in vascular cognitive disorder. (Under review).

Welton T, Ather S, Proudlock FA, Gottlob I & Dineen RA, 2016. Altered whole-brain connectivity in albinism. *Human Brain Mapping* 38(2):740-752.

Welton T, Evett L, Sherkat N & Brown DJ, 2016. A brain-computer interface for the dasher text entry system. *Universal Access in the Information Society* 15:77-83.

Welton T, Kent D, Auer DP & Dineen RA, 2015. Reproducibility of brain network metrics: a systematic review. *Brain Connectivity* 5(4):193-202.

Welton T, Kent D, Auer DP, Constantinescu CS & Dineen RA, 2014. Functionally relevant white matter degradation in multiple sclerosis: a tract-based spatial meta-analysis. *Radiology* 275(1):89-96.

Conference Oral Presentations

Welton T, Elsarraj AE, Auer DP & Dineen RA. Brain network organisation corresponds to changes in white matter integrity in MS. Oral presentation at: MS Frontiers; June 2015; London, UK.

Welton T, Auer DP & Dineen RA. Inter-scanner reliability of graph-theoretic brain network metrics. Oral presentation at: 23rd Annual Meeting of the ISMRM; May 2015; Toronto, Canada (Magna Cum Laude Award).

Conference Poster Presentations

Welton T, Auer DP, Constantinescu CS & Dineen RA. Reproducibility of Brain Network Metrics in People with Multiple Sclerosis. E-poster presentation at: ISMRM; April 2017; Honolulu, HI.

Welton T, Constantinescu CS, Auer DP & Dineen RA. Functional Brain Network Organisation is Related to Cognition and Fatigue in MS. Poster presentation at: ECTRIMS; June 2016; London, UK.

Welton T, Constantinescu CS, Auer DP & Dineen RA. Network dynamics during a sustained attention task in multiple sclerosis. Poster presentation at: 22nd Annual Meeting of the OHBM; June 2016; Geneva, Switzerland.

Welton T, Mohammad S, Proudlock F, Gottlob I & Dineen RA. Disrupted brain network topology in albinism. Poster presentation at: 20th Annual Scientific Meeting of the British Chapter of the ISMRM; September 2014; Edinburgh, UK.

Welton T, Auer DP & Dineen RA. Functionally-relevant white matter degradation in multiple sclerosis: a tract-based spatial meta-analysis. Poster presentation at: 22nd Joint Annual Meeting of the ISMRM-ESMRMB; May 2014; Milan, Italy.

Declaration

Except where acknowledged, I declare that this thesis is entirely my own work and is based upon research carried out in Radiological Sciences, Division of Clinical Neuroscience, University of Nottingham, between October 2013 and April 2017. The material contained herein has not been presented, or is currently being presented, either wholly or in part for any other degree or qualification.

Thomas Welton

Acknowledgements

I would like to thank the following people:

- My primary supervisor, Dr Rob Dineen, for his guidance and support in all aspects of my PhD studies.
- Co-supervisors, Prof Dorothee Auer and Prof Cris Constantinescu, who helped conceive the project and plan the experiments.
- The UK MS Society and its benefactors, who provided the primary funding for the project.
- The students and staff of the Radiological Sciences research group (Sir Peter Mansfield Imaging Centre) for their friendship and support. In particular, I would like to acknowledge the help of my colleagues Dr Afaf Elsarraj, Dr Daniel Kent and Dr Yue Xing with whom I collaborated most.
- Prof Nadina Lincoln, Dr Roshan das Nair, Ms Cara Knight and Ms Holly Chappell for their help with recruitment and collaboration with the CRAMMS trial.
- Dr Sarim Ather, Dr Frank Proudlock and Prof Irene Gottlob of the Ulverscroft Eye Unit, Leicester, with whom I collaborated.
- The radiographers at the Medical Imaging Unit in the Queen's Medical Centre: Ms Anita French, Dr Brett Haywood and Mr Andrew Cooper who operated the MRI scanner.

- MS-Research, the University of Nottingham Graduate School, the International Society for Magnetic Resonance in Medicine and the Guarantors of Brain for additional travel funding.
- All of the contributors of data to the meta-analysis chapter: Dr Astrid Blaschek (Ludwig-Maximilian University of Munich, Germany), Dr Christian Enzinger (Medical University of Graz, Austria), Prof Franz Fazekas (Medical University of Graz, Austria), Dr John D Fisk (Dalhousie University, Canada), Dr Antonio Giorgio (University of Siena, Italy), Ms Hanneke E Hulst (VU University Medical Center, the Netherlands), Prof Heidi Johansen-Berg (Oxford University FMRIB Centre, United Kingdom), Dr Daniel Keeser (Ludwig-Maximilian University of Munich, Germany), Mr Kyle C Kern (University of Miami Miller School of Medicine, FL, USA), Prof Lauren B Krupp (Stony Brook Medicine, NY, USA), Dr Yaou Liu (Capital Medical University, China), Dr Sara Llufriu (Hospital Clinic of Barcelona, Spain), Mr Eloy Martinez-Heras (Hospital Clinic of Barcelona, Spain), Dr Erin L Mazerolle (University of Calgary, Canada), Dr Mihaela Onu (Theodor Burghele Clinical Hospital, Romania), Dr Albert Saiz (Hospital Clinic of Barcelona, Spain), Dr Menno M Schoonheim (VU University Medical Center, the Netherlands), Dr Nancy L Sicotte (Cedars-Sinai Medical Center, CA, USA), Prof Stephen M Smith (Oxford University FMRIB Centre, United Kingdom) and Dr Mark E Wagshul (Stony Brook University and Albert Einstein College of Medicine, NY, USA).
- The volunteers willing to be scanned and tested for the research.

- Data used for the inter-scanner reproducibility experiment (section 6.5) were downloaded from the Function BIRN Data Repository (http://fbirnbdr.birncommunity.org:8080/BDR/), supported by grants to the Function BIRN (U24-RR021992) Testbed funded by the National Center for Research Resources at the National Institutes of Health, U.S.A.

List of Contents

Abstract i
List of Publications October 2013 – April 2017 v
Declaration vii
Acknowledgements viii
List of Contents xi
Part I: Introduction and Background 1
1. Introduction 2
1.1. Background to the Problem 2
1.2. Problem Statement 6
1.3. Aims and Objectives 6
1.4. Thesis Structure 8
2. Multiple Sclerosis as a Disconnection Syndrome 10
2.1. Neurobiology 10
2.2. Pathology 11
2.3. Clinical and Radiological Presentation 15
2.4. Disconnection Syndromes 24
3. Imaging the Multiple Sclerosis Brain 26
3.1. The Magnetic Resonance Signal 26
3.2. Imaging MS in Clinical Practice 31
3.3. Imaging MS in Research 32
4. The Graph-Theoretic Model of Brain Networks 35
4.1. Graph Theory 36
4.2. Networks 38
4.3. Brain Networks Relevant to Cognition and Fatigue in MS 41
Part II: Research Studies 52
5. The Relevance of White Matter Degradation to Impaired Function in Multiple Sclerosis 53
5.1. Introduction 53
5.2. Methods 55
5.3. Results 60
5.4. Discussion 75
6. Reliability of Brain Network Properties 79
6.1. Introduction 79

6.2. Aims and Hypotheses 84
6.3 Systematic Review of Reproducibility of Brain Network Metrics 85
6.4. Analysis of Reproducibility of Graph Metrics in People with MS 103
6.5. Analysis of Reliability of Graph Metrics between Scanners 115
6.6. Discussion 121
7. Relationships between Cognitive Impairment and Brain Network Properties in Multiple Sclerosis 131
7.1. Introduction 131
7.2. Generic Methods and Sample Characteristics of the BraNDy-MS Cohort 132
7.3. Methods 160
7.4. Results 165
7.5. Discussion 177
8. Responsiveness of Brain Network Properties 182
8.1. Introduction 182
8.2. Methods 184
8.3. Results 185
8.4. Discussion 194
Part III: Discussion 198
9. Discussion and Conclusions 199
9.1. Summary of Findings 199
9.2. Are Graph-Theoretic Brain Network Metrics Good Candidate Markers of Cognition or Fatigue in MS? 201
9.3. Limitations 204
9.4. Future Directions 206
9.5. Conclusions 207
Appendices 209
A. Retrieving local maxima from TBSS t-statistic maps 209
B. Research Ethics Committee Approval Letter 210
C. Consent Form 211
D. Information Sheet 212
References 214

Part I: Introduction and Background

1. Introduction

1.1. Background to the Problem

Multiple sclerosis (MS) is a disease of the central nervous system (CNS) affecting an estimated 127,000 people in the UK [1] and 2.5 million people worldwide [2]. In over half of those cases, MS manifests, among a spectrum of other neurological symptoms, as a progressive decline in one or more cognitive faculties and, even more frequently, as chronic fatigue [3, 4]. These conditions are highly disabling: they affect quality of life and the ability of the individual to function normally in society; for example, 70-80% of people with MS are unemployed [5, 6]. As a result, MS has a large economic impact: one study ranked MS 14th in financial burden of brain disorders to the UK, costing an estimated £2.3 billion annually, ahead of Parkinson's disease (£2.0 billion), and brain tumours (£0.7 billion) [7]. Other estimates range from £2.6-3.4 billion annually [8, 9].

Multiple sclerosis, therefore, presents important economic, social and humanitarian challenges. A significant portion of such challenges are related to the presence of cognitive impairment and increased susceptibility to fatigue. These are among the most common symptoms of MS: the most comprehensive study of the natural history of symptoms in MS to date performed by Kister et al [10] in a cohort of 36,000 people with MS showed that, although prevalence and severity of symptoms in MS vary considerably, at diagnosis, 26% of people with MS reported moderate or severe cognitive impairment, which increased to 48% at 10 years' disease duration. 56%

reported moderate or severe fatigue at diagnosis, rising to 79% after 10 years disease duration. Other estimates for the prevalence of cognitive impairment and fatigue range from 43-70% [3, 11, 12] and 50-66%, respectively [13-15]. The range and severity of deficits can vary widely between patients: various aspects of cognition can be affected, including attention, information processing efficiency and speed, executive function and long-term memory. Processing speed, visual learning and memory are the most commonly degraded aspects of cognition, while other aspects tend not to be affected; for example, short-term memory and verbal skills [16].

Due to their high impact, a major research effort has been given to developing and testing potential treatments for these symptoms. Trials of symptomatic treatments specifically for cognitive symptoms have been relatively few, underpowered, and their results equivocal [17]. A Cochrane review of studies of pharmacological treatments for memory impairment in MS found no convincing evidence for or against pharmacological treatments but concluded that better methodological quality was needed [18]. A review of trials of non-pharmacological cognitive rehabilitation concluded that the results of such studies to date were heterogeneous [19].

A recurrent problem in studies of this kind is that cognitive impairment and fatigue are difficult to measure in a way that is objective, specific and reliable. Most often, cognitive impairment is assessed by neuropsychological tests or questionnaires such as the Selective Reminding Test, the California Verbal Learning Test, the Symbol Digit Modalities Test (SDMT) the Paced Auditory Serial Addition Test (PASAT) [18] and fatigue by a Visual Analogue Scale or questionnaires like the MS-Specific Fatigue Scale, Fatigue Severity Scale or

the Neurological Fatigue Index (NFI-MS) [20]. There are a number of disadvantages to this approach. First, self-reported cognitive dysfunction and performance on neuropsychological tests have a poor correspondence [21]. In addition, while scales like the SDMT are quick to administer and can be internally reliable, their correlation with gold standard tests is poor [22]. A likely contributing factor to this is that responses to neuropsychological tests and questionnaires are dependent on the individual's interpretation of the question and their current state psychological state, e.g. alertness or mood. Each of these tests is broad in scope and measures multiple aspects of cognition. For example, the SDMT assesses information processing speed and flexibility, immediate recall, the ability to perform calculations, the ability to vocalise the response and the ability to hear and understand the instructions and the test numbers. This is a problem because it means that such tests are, to an extent, nonspecific, providing a summary score which may not be representative.

Magnetic resonance imaging (MRI) may provide a way to measure the neurobiological correlates of cognitive impairment and susceptibility to fatigue. These correlates based on MRI data might be assessable more objectively, more reliably and with greater specificity. Test-retest reliability of tests of cognitive impairment and fatigue is adequate, but can be susceptible to practice effects [23-26]. MRI measurements require no input from the subject and can provide detailed spatial information about observed brain abnormalities. Where the outcomes of questionnaires are summary measures of multiple aspects of cognition and affected by many factors, MRI measurements have the ability to be specific to a particular feature of brain

health. However, conventional imaging markers like total white matter lesion load (TWMLL) and atrophy do not sufficiently explain clinical status [27, 28]. For example, correlations between lesion load and clinical disability have ranged from $\rho = 0.09$ to $\rho = 0.60$, often falling at the lower end of this range [29].

Developments in MRI technology and image analysis techniques have opened the door to new potential types of imaging markers. This is recognised by leading academics and charities engaged in funding MS research; for example, the recently-published research strategies for the MS society and the International Collaborative on Progressive MS, which stress the importance of developing imaging markers of cognitive impairment [30, 31]. One such new development in brain imaging research currently is the application of graph theory to the modelling of brain networks using MRI data [32]. Graph theory shows promise as an holistic model of brain connectivity in neurological disease [33], and may be particularly important for understanding brain network changes relevant to cognition in MS [34]. In graph-theoretic analysis of the brain, a graph consists of nodes representing brain regions or voxels, and edges between the nodes representing some measure of connectivity. From this structure, summary descriptors of the graph's organisation can be extracted. The idea that graphs could be a useful model of the MS brain is supported by a growing body of evidence which suggests that symptoms in MS arise as a result of disconnection of brain regions by damage to the interconnecting white matter (WM) tracts [35, 36]. This is known as the "disconnection hypothesis". Damage to cortical grey matter also occurs in MS and is relevant for brain network structure [37, 38].

1.2. Problem Statement

Cognitive impairment and fatigue are prevalent and impactful symptoms of multiple sclerosis. In order to accurately test the efficacy of treatments for these symptoms, effective markers are needed. Existing markers may be insufficient because, while they are related to clinical impairment, they are subjective and broad in scope. MRI-based markers have the potential to provide more detailed, objective, reliable and specific complimentary information to existing markers.

The general challenge faced by researchers is therefore to identify MRI-based markers of cognitive impairment and fatigue which are suitable for use in clinical studies. Markers based on the graph-theoretic model of brain connectivity may be particularly suited to this purpose because of the increasingly prevalent view that proper brain function is determined by the structure and intercommunication of grey matter networks. In addition, the graph-theoretic approach overcomes a common limitation of seed-based analyses which requires the a priori definition of a region of interest (ROI).

1.3. Aims and Objectives

Thus I was motivated to test the measurement of graph-theoretically derived properties of brain networks based on MRI data, with the view to evaluating their potential usefulness as candidate markers of cognitive impairment and fatigue in MS. I formed the following overarching hypothesis:

> *Summary descriptors of graph-theoretic brain network organisation are good candidate markers of cognition or fatigue in MS.*

The hypothesis is dependent upon two predicates:

1. The presence of cognitive symptoms in MS is attributable to the disruption of macroscopic brain networks by damage to the interconnecting white matter (the disconnection hypothesis).
2. Graph theory is an interpretable and generalisable way to model brain network organisation.

I test the hypothesis and predicates by asking three intermediate research questions:

1. *Are cognitive impairment and fatigue in MS related to disruption to macroscopic brain networks?* Answering this question constitutes testing predicate 1, the disconnection hypothesis, and would establish a basis for the study of cognitive phenomena in the context of network disruption.
2. *What are the limitations of graph-theoretic brain network models?* This question corresponds to predicate 2 and addresses the suitability of the graph-theoretic model to the problem.
3. *Are graph-theoretic brain network properties useful markers of cognitive impairment or increased fatigability in MS?* This question directly addresses the main hypothesis. Answering it involves evaluating candidate markers against a set of three *a priori* criteria (figure 1): firstly, they should measure what they purport to measure (validity); second, they should be statistically reliable across subjects, time and different settings and; third, they should be responsive to manipulation or change in the subject in the immediate term.

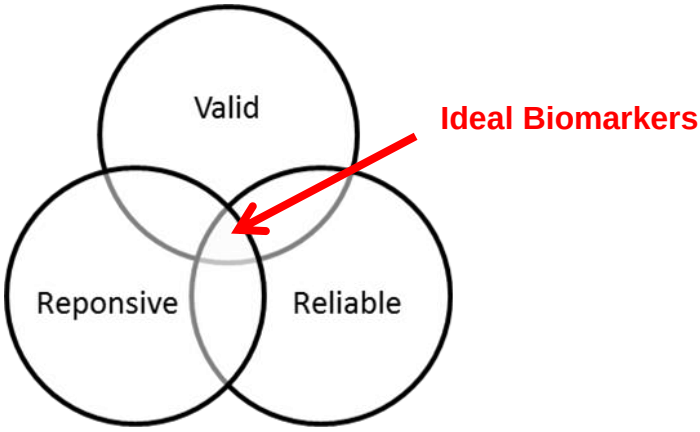


Figure 1. Venn diagram showing the proposed criteria for evaluating biomarkers. The ideal biomarker meets all three criteria: reliability, validity and responsiveness. Alongside these are other practical criteria such as ease of interpretation, time-efficiency and cost-efficiency. These key characteristics were chosen based on the recommendations of the NHS Health Technology Assessment Programme [39] and on multiple reviews of the assessment of outcome measures in clinical trials [40-42].

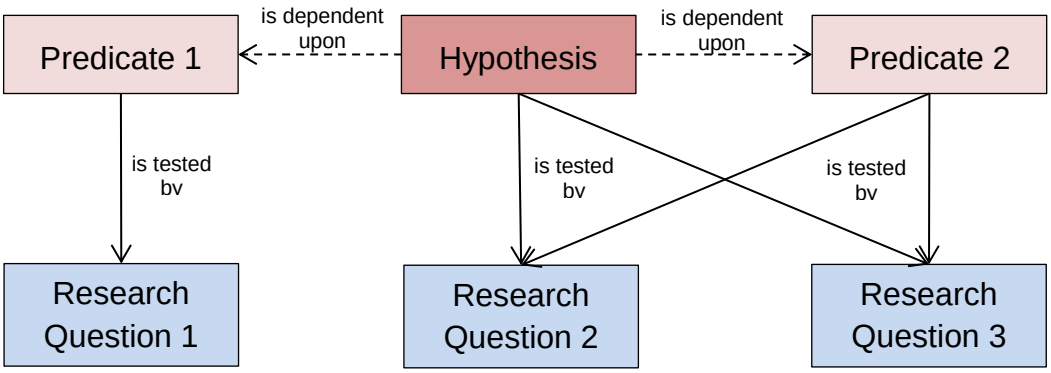


Figure 2. Diagram showing that the hypothesis is dependent on the two predicates and which of the three research questions aim to test which.

1.4. Thesis Structure

The thesis is organised into 3 parts: Part I includes the introductory and background information chapters, Part II includes descriptions of studies and results and Part III includes a discussion and the conclusions.

Part I

In this chapter, I have given the broad context and described my motivations and rationale. I have laid-out the nature of the problem and how I proposed to tackle it. Chapter 2 gives background information on multiple sclerosis and introduces the idea of a disconnection syndrome. Chapter 3 describes how MS is imaged in clinical practice and the relevant techniques being applied in research. Chapter 4 describes the mathematics of graphs and how graphs are being applied in the study of the brain, specifically in MS.

Part II

Chapter 5 provides evidence for research question 1 through meta-analysis of data from previous studies of white matter damage in MS and its clinical and cognitive relevance. Chapters 6, 7 and 8 address the requirements for evidence of validity, reliability and responsiveness, respectively, to address research questions 2 and 3. Chapter 7 includes descriptions of the protocols for recruitment, imaging and cognitive testing for the “Brain Network Dysfunction in MS” (BraNDy-MS) cohort, characterisation of the cohort as well as methods used in the subsequent chapters.

Part III

Chapter 9 is a discussion of how the results of the experiments described in this thesis come together to answer the main research question. It includes a discussion of the limitations, possible future directions and conclusions.

2. Multiple Sclerosis as a Disconnection Syndrome

First named in 1868 by Jean-Martin Charcot [43], but described as early as the middle ages, MS was among the first diseases to be studied scientifically. Despite a long history of research into the causes of and potential treatments for MS, it remains incurable. In this chapter, I describe the pathology and diagnosis of MS and focus on the disconnection hypothesis, a theory about how symptoms arise in MS. While many of the facts given here are common knowledge, I generally refer to the textbook of Compston, McDonald [44].

2.1. Neurobiology

In order to understand the pathology of MS, it is useful to first describe some basic neurobiology. The CNS is composed of two main types of cell: neurons and glia. Neurons process and transmit electrical and chemical information to one another in the CNS, and to the various other systems of the body. Each neuron is composed of a soma (main body), dendrites which branch out from the cell body and an axon which is a single long extension from the soma. Neurons are able to transmit electrical information by maintaining a voltage gradient between the inside and outside of the cell, which is a result of differences in the concentration of various ions across the cell membrane. Large enough differences in concentrations can reach the threshold to trigger an “action potential”, at which the cell emits an electrochemical pulse along its axon. Glia (from the Greek for “glue”) are responsible for supporting and

protecting the neurons. One of their most important functions is the formation of myelin. Myelin, which is a good electrical insulator, surrounds axons, forming a sheath. This sheath helps electrical signals propagate along the length of the axon faster and with less attenuation. The following sections describe in steps how, in MS, the normal functioning of the CNS is disrupted (figure 3).

2.2. Pathology

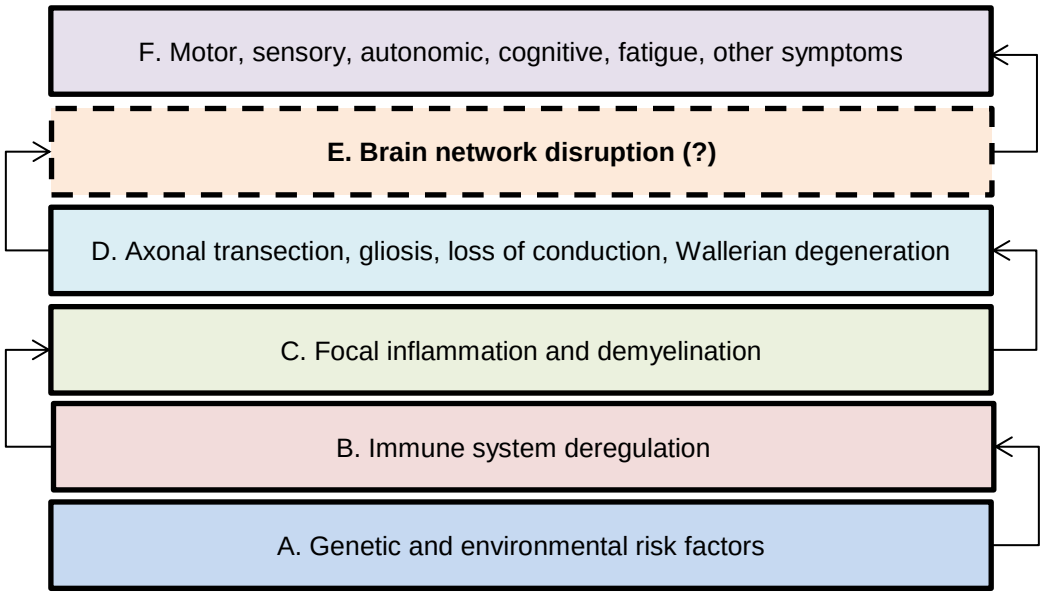


Figure 3. Basic overview of MS pathology and role of brain network disruption according to the disconnection hypothesis. Part E (dotted outline) represents a missing link: disruption to brain networks due to damage to the interconnecting white matter. This connection provides an explanation for why neurodegeneration results in specific symptoms.

Susceptibility to and severity of MS symptoms are determined by both heritable and environmental factors (figure 3A). The human leukocyte antigen has the largest genetic contribution to MS risk [45]. The human leukocyte antigen complex is partly responsible for the way in which the immune system

identifies exogenous proteins; specifically, the DRB1 ($r=-0.702$, $p=3.9\times10^{-10}$), DRB5 ($r=-0.728$, $p=4.4\times10^{-11}$) and DQB1 ($r=-0.647$, $p=2.3\times10^{-8}$) genes are strongly associated with susceptibility to MS [46]. Mutations in a second gene, IL7R, are also associated with MS (OR=1.15, 95% CI=1.06-1.24, p=0.0009) [47]. The IL7R gene is responsible for the production of two receptor proteins in immune cells. That both genes are related to immune function suggests that they are involved in an autoimmune response that damages the myelin in MS. However, the precise details of how this occurs are not known. A third gene, CYP27B1, which encodes a vitamin D-activating enzyme, may also play a role in MS risk (p=0.00001) [48], and may provide an explanation for one of the strongest environmental risk factors: hypovitaminosis-D. Studies of twin pairs and familial aggregation show that risk increases with relatedness [49].

One strong environmental risk factor for MS is infection with the Epstein-Barr virus, which has a relative risk of up to 3.9 in patients with antibodies to EBNA-2 [50]. Incidence increases rapidly after 18 years old and is greater in women than in men (with female-to-male ratios between 1.5 and 2.5) [51]. Incidence increases with latitude, and is low in Asia and in areas with a tropical climate [52], although this effect appears to be lessening in recent times [53]. More recently, sodium intake was found to be associated with disease activity in MS, through the autoimmune differentiation of exo- and endo-genous cells: individuals on a high-sodium diet were 3.4 times more likely to develop a new lesion [54]. While nearly all people with MS will experience fatigue, cognitive impairment is more likely to occur in those with low premorbid intelligence [55]. Another major risk factor for MS is smoking,

with a large meta-analysis reporting a risk ratio of 1.48 (95% CI 1.35-1.63) [56]. Additionally, people with MS who smoke tend to be heavier smokers than people who smoke and do not have MS [57]. For many of these factors, a precise explanation for why they lead to the following described steps is currently lacking.

It is believed that the disease process begins with the increased infiltration of autoreactive T-cells across the blood-brain barrier (figure 3B). Usually, regulatory T-cells (specifically CD4+CD25+) would supress autoreactive T-cells (faulty T-cells which mistakenly react to the body, rather than pathogens) but, in MS, this process is disrupted [58]. Apoptosis of the T-cells is prevented by the overexpression of a protein (ARRB1) which dampens the cellular response to stimuli [59]. The result is inflammation of the affected tissues (figure 3C, figure 4). Inflammation is the movement of plasma and leukocytes from the blood into injured tissue. In MS, inflammatory demyelination occurs in focal lesions which are usually located in white matter tracts and diffusely throughout normal-appearing white matter (NAWM).



Figure 4. An active multiple sclerosis lesion stained for myelin (black). The arrowheads point to myelin debris inside the demyelinated area. The scale bar represents 400 μm. Image copied from Trapp, Peterson [60].

Demyelination eventually leads to damage to neurons via a number of proposed routes, which include microglia activation, chronic oxidative injury and mitochondrial damage in axons (figure 3D, figure 5) [61]. Transected or demyelinated axons cannot transmit impulses at the normal velocity, may exhibit ephaptic coupling more frequently and may be affected by Wallerian degeneration, trans-synaptic degeneration or other neurodegenerative processes [62]. These processes may be mitigated by anti-inflammatory cytokines and by remyelination.

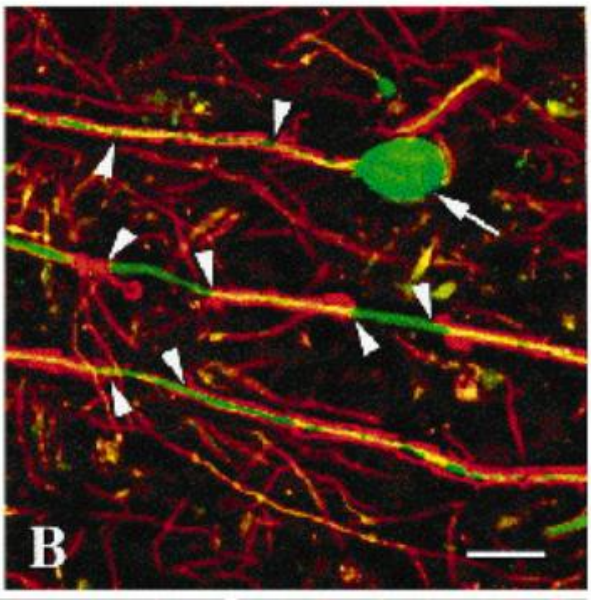


Figure 5. Confocal image of demyelination in MS from Trapp, Peterson [60]. The image is from the edge of an active lesion. Red indicates myelin, green indicates the axon.

Traditionally, it has been assumed that the ensuing symptoms are a direct result of neurodegeneration (figure 3F). More recently, however, an intermediate step, network disruption, is thought to better explain the presence of symptoms in MS. The next sections describe the typical clinical and radiological presentation of MS, and the disconnection hypothesis, which seeks to explain those symptoms in the context of brain network disruption.

2.3. Clinical and Radiological Presentation

No single test can definitively diagnose MS. Instead, neurologists apply sets of criteria which incorporate evidence from multiple sources. The most commonly used criteria for diagnosis of MS are called the McDonald criteria [63], which require dissemination in both time and space. The McDonald criteria are summarised in table 1. Further updated details and

recommendations were made by the MAGNIMS study group in 2016 based on new MRI technology and data for the dissemination in time and space [64].

Table 1. Summary of the 2010 updated McDonald criteria, from Polman, Reingold [65].

Clinical Presentation	Additional Data Needed for MS diagnosis
≥ 2 attacks; ≥ 2 lesions	None
≥ 2 attacks; 1 lesion	DIS [a]
1 attack; ≥ 2 lesions	DIT [b]
1 attack; 1 lesion	DIS [a] and DIT [b]
Insidious neurological progression suggestive of MS (PPMS)	1 year of disease progression plus 2 of 3: positive CSF; DIS based on ≥ 1 T2 lesion in the brain; DIS based on ≥ 2 T2 lesions in the spinal cord

[a] DIS: Dissemination in space. Can be demonstrated by: ≥ 1 T2 lesion in at least 2 of 4 MS-typical regions (periventricular, juxtacortical, infratentorial or spinal cord); or a further clinical attack which implicates a different CNS site.

[b] DIT: Dissemination in time. Can be demonstrated by: simultaneous asymptomatic gadolinium-enhancing lesion and non-enhancing lesion; or a new lesion on a follow-up MRI; or await a further clinical attack.

Clinical presentations can vary widely and often cannot be linked to radiological signs. Similarly, radiological signs cannot always be linked to clinical presentation. This is known as the clinico-radiological paradox [27]. Some of the most common signs and symptoms are: optic neuritis, fatigue, cognitive impairment, sensory symptoms, cerebellar symptoms, nystagmus, motor symptoms, sphincter disturbance and chronic pain. Lesions at some CNS sites do have clear correlations to clinical signs and symptoms; for example, lesions in the spinal cord are most commonly associated with weakness, stiffness, spasms, bladder dysfunction, constipation, erectile dysfunction or upper motor neuron signs whereas lesions in the cerebrum are more commonly associated with sensory deficits, motor deficits and cognitive impairment and anterior optic pathway lesions with painful vision loss. The

most commonly affected CNS sites are the optic nerves and chiasm, the cervical cord and brainstem. In the cerebrum, MS lesions tend to appear around the ventricles, infratentorially or juxtacortically. Some typical appearances of lesions in MS are shown in figure 6.

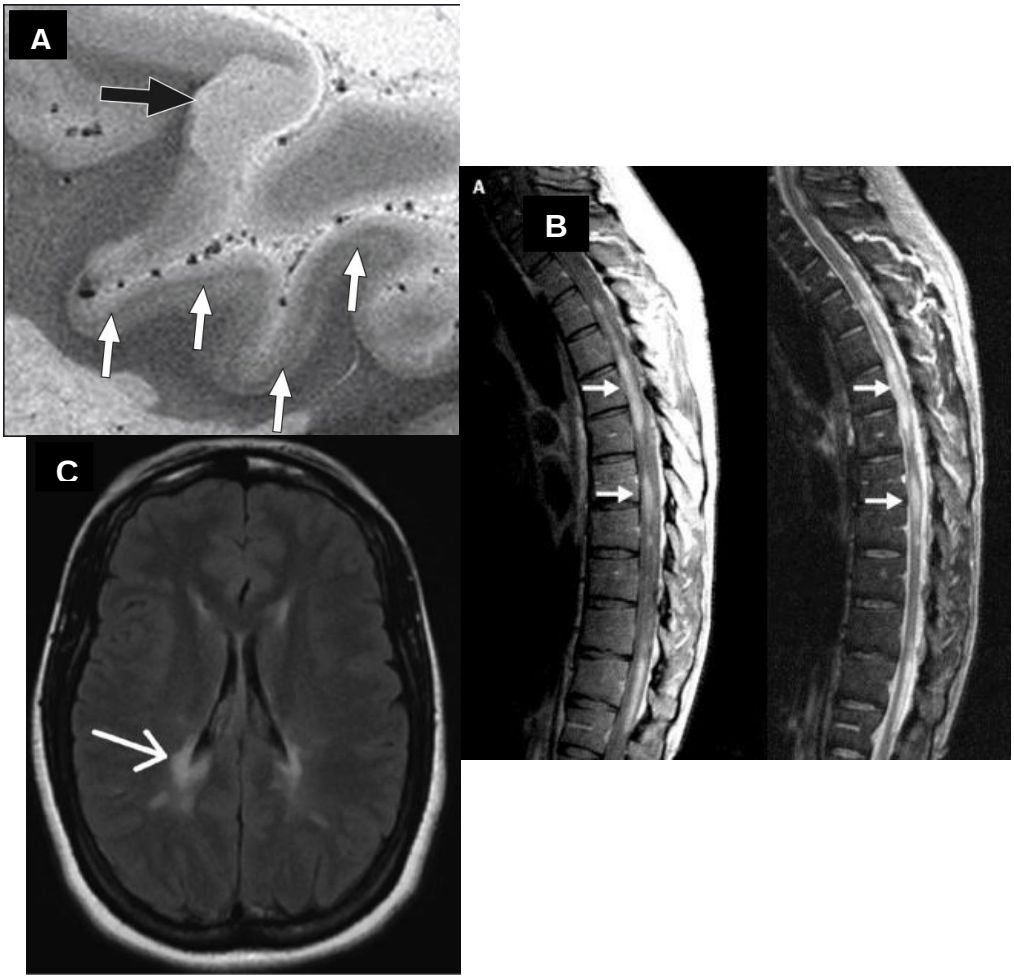


Figure 6. Radiological presentation of MS. (A) T2*-weighted image at 7T of a lesion in the cortex (black arrow) and the line of Gennari visible (white arrows), from Pitt, Boster [66]. (B) Sagittal proton-density and T2-weighted images at 7T of lesions in the spinal cord with swelling, from Lycklama, Thompson [67]. (C) Periventricular lesions (arrow) imaged at 3T with a FLAIR sequence, from Klawiter [68].

MS cases are generally classified based on the pattern of severity and relapses over time into four types: relapsing-remitting (RR), secondary

progressive (SP), primary progressive (PP) and progressive relapsing (PR; figure 7). Clinically-isolated syndrome (CIS) is the name given to the situation where the patient has had the first neurological episode but there is not yet sufficient evidence for a definite MS diagnosis. Likewise, a radiologically-isolated syndrome (RIS) is the situation where a brain MRI shows lesions suggestive of MS but no clinical symptoms are present. Most MS patients begin with a CIS and later convert to MS. The risk of conversion from CIS to MS within 10 years is 60-80% in patients who have asymptomatic lesions and 20% in patients who do not [69].

Approximately 85% of MS patients have the RR subtype at diagnosis, which is characterised by recurrent episodes with a return to baseline in between. Most patients with RRMS will progress to SPMS, but some (approximately 15-25%) will not, and will have a benign form of MS. The frequency of relapses increases over time, being approximately yearly in the early stages. As the frequency increases, the recovery between relapses may be incomplete. Incomplete remissions and a predominance of motor symptoms are associated with worse prognoses [70].

In SPMS, exacerbations are followed by a progressive worsening of symptoms with time. SPMS typically comes after a period of 3 years after the patient being classified as RRMS, and in 90% of patients within 25 years. This change represents the point where the brain starts to lack the capacity to compensate for the injury, either by remyelination or redundancy.

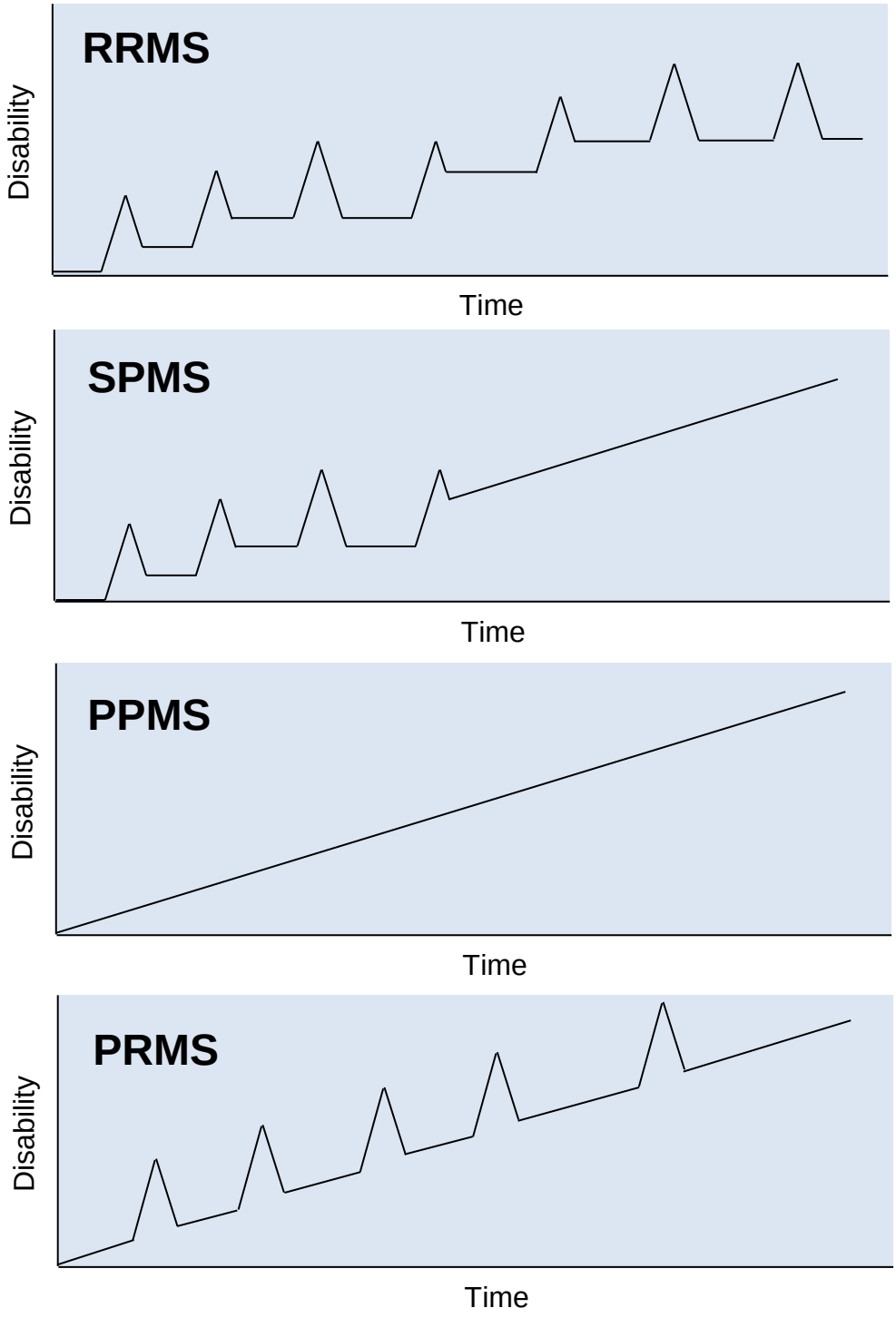


Figure 7. Classification of MS subtypes. RRMS is characterised by unpredictable clinical attacks and remissions which may or may not leave permanent deficits. SPMS is the name given to the disease when the patient begins with RRMS but then changes to a steady increase in disability without relapses. PPMS is a gradual progression in disability without relapses or remissions from the disease onset. PRMS is a progressively increasing disability with relapses super-imposed.

Cognitive Impairment

Cognitive impairment is one of the major factors related to disability in MS. Estimates for prevalence range from 40-60% [3, 12, 71, 72]. Cognitive impairment can occur independently from other symptoms and is only weakly correlated with disease duration and physical disability, which can make its identification and measurement more difficult [73, 74]. While verbal skills and fluency are generally preserved, several other areas of cognition can be affected, most often including visual memory and information processing speed (51.9% and 54.3%, respectively) [3]. General intelligence is not greatly altered in MS [75]. The usual presentation of cognitive impairment in MS, then, is of subtle impairments in specific cognitive domains which can differ greatly between individuals.

There are a number of factors with known moderating effects on cognitive impairment in MS. One major factor is MS subtype, with people in the SP stage tending to have worse cognitive performance compared to both RR and PP subtypes, although it is not clear why this is the case [76-79]. A consistent positive link between depression and cognitive impairment has also been demonstrated in MS [80]. Depression is common symptom of MS (present in up to 60% of people [81]) and, while its cause is not clear, it is associated more with cerebral lesions than spinal cord lesions [82] and is associated with hypoconnectivity between cortical and subcortical brain networks [83, 84].

Cognitive impairment is highly impactful on the daily lives of people who are affected. For example, measures of quality of life are closely negatively correlated with cognitive impairment [85] in people with MS, as well as with MS diagnosis in general [86], depression [87], disability [88] and disease

progression [89]. Cognitive impairment is detrimental to productivity [90] and leads to unemployment for up to 80% of individuals [91, 92]. This is especially damaging due to the typical onset of disease being during the period of life when individuals are most active (between the ages of 20 and 40). In comparison, physical disability accounts for far less variance in employment status (approximately 14% [93]) than does cognitive impairment (approximately 49% [90]). Measures of processing speed, verbal memory and executive function in particular are related to employment status [12]. Cognitive impairment also affects social activity: cognitively impaired people with MS were engaged in fewer social activities than cognitively preserved controls [92].

Treatments for cognitive impairment take one of two forms: pharmacological interventions or rehabilitation programmes. Of the pharmacological interventions, two main approaches can be identified: disease modifying therapies (DMTs) and drugs targeted specifically at improving cognition. DMTs are drugs which can reduce the frequency and severity of relapses and slow the rate of white matter damage. In trials testing the efficacy of DMTs, cognitive impairment has not been a focus of the study, and was only included as a secondary outcome. As such, it difficult to determine the effect of DMTs on cognitive impairment. However, some positive effects have been found for interferon beta-1a and interferon beta-1b [94, 95]. The non-DMT pharmacological interventions receiving the most attention are acetylcholinesterase inhibitors like donepezil, which show significant improvements in memory performance [96], but this research is still preliminary [97].

Findings from trials of cognitive rehabilitation in MS have been inconclusive. Such interventions involve a “structured, planned experience derived from an understanding of brain function which ameliorates dysfunctional cognitive and brain processes caused by disease or injury and improves everyday life function” [98]. Repeated stimulation of brain areas relevant to memory, for example with a computerised practice exercise, is thought to improve memory function by means of neuroplasticity [99]. Other forms of rehabilitation might involve behavioural strategies enacted in a group, simply providing education and awareness of cognitive problems or teaching practical tools such as cognitive strategies, list-making or diary-writing. A 2016 Cochrane review of memory rehabilitations in MS combined evidence from 15 randomised controlled trials and found limited evidence for the effectiveness of memory rehabilitation on both memory performance and measures of real-world functional ability, with significant effects for intervention on general memory performance in both short and long term (standardised mean differences 0.23, 95% CI 0.02 to 0.41 and 0.26, 95% CI 0.03 to 0.49, respectively) [100]. The authors concluded that further randomised controlled trials of high methodological and reporting quality should be performed using more ecologically-valid outcomes. A systematic review identified 33 studies of cognitive rehabilitation in MS and found that, although most studies reported overall improvements in cognitive performance, no definitive conclusions could be drawn [101].

Cognitive impairment is usually assessed using neuropsychological tests or groups of tests. The most common test for cognitive impairment applied in MS research is the PASAT, which involves mentally adding the last two numbers

of a sequence read out by a recording and saying the answer. The test requires fast processing of auditory information, working memory, attention and flexibility. The most common neurological test battery for function in MS, the Multiple Sclerosis Functional Composite (MSFC), uses the PASAT as a measure of cognition as its third component. Other frequently-used groups of tests for cognition in MS include the Brief Repeatable Battery of Neuropsychological Tests [102] and the Brief Cognitive Assessment for Multiple Sclerosis [103]. Other tests commonly included in these batteries are the SDMT [104], the Selective Reminding Test [105] and the California Verbal Learning Test [106], all of which are validated in samples of people with MS.

Fatigue

Fatigue is the most common symptom of MS, with a prevalence of at least 75% [107], and it has been described as the most disabling symptom by 80-85% [108, 109] of individuals with MS. It is related to reductions in quality of life [110], physical activity [111] and depression [112]. The cause for increased levels of fatigue in people with MS is not known, but it may have both primary (CNS damage, altered patterns of cortical activity) and secondary (sleep disturbance, depression) mechanisms [113]. As fatigue is an ambiguous and multidimensional symptom, some discussion has been made on creating a single correct definition. A proposed taxonomy distinguishes the subjective perception of fatigue and objective measures of “fatigability” as measured by changes in task performance [114]. Fatigue is usually measured by self-reported questionnaires, like the Modified Fatigue Impact Scale (MFIS) [115] or Fatigue Severity Scale [116]. While there are limitations to these tests, such as poor correspondence to functional status

[117] and muscle fatigue [118], they are very widely-used in multiple sclerosis clinical assessment and research [119].

2.4. Disconnection Syndromes

A disconnection syndrome is one whose symptoms are explained by disconnection of distal brain regions by damage to the interconnecting white matter. The idea was laid-out by Norman Geschwind in 1965:

> “lesions of association cortex, if extensive enough, act to disconnect primary receptive or motor areas from other regions of the cortex in the same or in the opposite hemisphere. ... Thus a ‘disconnexion lesion’ will be a large lesion either of association cortex or of the white matter leading from this association cortex." [120]

Prior to Geschwind, research in the late 1800s contained descriptions of what are now called disconnection syndromes: conduction aphasia, visual agnosia, apraxia and pure alexia [121].

In recent years, cognitive impairment in multiple sclerosis has also been described as a “multiple disconnection syndrome” [35, 122], referring to the fact that multiple cognitive domains can be affected. Evidence for this can be found in studies of white matter tract integrity and its correlation with clinical and cognitive measures, of which there are many (reviewed in chapter 5). In multiple sclerosis, the effect is arguably much more complex than in the classical disconnection syndromes: dysfunctional pathways are not necessarily completely transected and may instead be affected by demyelination or Wallerian degeneration to varying degrees (in which an axon

is cut or crushed and the distal part of the axon degenerates). The distribution and severity of disconnections can vary widely and the damage to white matter in MS occurs both focally in lesions and diffusely, even in NAWM. Viewing MS-related cognitive impairment from this perspective has led to special interest in “connectomic”-type analysis methods, which aim to produce and study maps of connections within the nervous system [123]. These methods could provide a convenient way to measure the impact of multiple disconnections on brain network structure and function.

3. Imaging the Multiple Sclerosis Brain

3.1. The Magnetic Resonance Signal

The magnetic resonance signal is fundamentally based on the magnet-like behaviour of atoms caused by their spinning motion. When placed in a strong magnetic field, atoms become aligned with its flux. A hydrogen atom with a single proton in its nucleus, ^{1}H, spins at a rate of 127.728 MHz in a magnetic field of 3 Tesla (which is a standard field strength in modern clinical MRI scanners). The rate of spin changes proportional to a constant, γ, the gyromagnetic ratio, which is specific to each type of atom. This relationship is described by the Larmor equation:

$$\omega = \gamma B_0$$

Where ω represents the Larmor frequency (or spin rate of the atom), γ is the gyromagnetic ratio and B_0 is the applied magnetic field strength. For a ^{1}H nucleus, γ is 42.576 MHz/Tesla.

A second important concept is that of precession: as an atom spins about its axis in a strong magnetic field, the axis itself is never fixed, causing the motion to resemble that of a spinning top. When a radiofrequency pulse at the Larmor frequency of an atom is applied, two simultaneous processes occur: (1) that atom becomes excited and the axis about which it precesses moves away from alignment with the magnetic field and (2) the atom becomes in-phase with others of its type; i.e., they begin to precess in synchrony.

When the radiofrequency pulse has ended, the atoms' alignment and phase return to their baseline state, emitting radio waves as they do so. As atoms return to their baseline state, two simultaneous processes occur: (1) they return to their original alignment with the magnetic field and (2) they fall back out of phase with each other. The time it takes for these processes to occur, known as T1 and T2, respectively, govern the radiofrequency signal which is detected and subsequently used to construct an image. T1 refers to process (1) and represents the time taken for atoms to return to alignment (longitudinal relaxation). T2 refers to process (2) and represents the time taken for atoms to become dephased (transverse relaxation). A third decay rate, known as T2*, is similar to T2 but takes into account inhomogeneity in the magnetic field caused by differences in the composition of the surrounding tissue. Different tissue qualities produce different T1 and T2 times in the emitted signal, allowing differentiation between tissue types. The timings and order of the radiofrequency pulse sequence can be altered to weight the signal toward either process.

Another concept to consider is that of spatially encoding the MRI signal. To distinguish where in the brain a particular signal came from, each part is assigned a different frequency using gradient fields. These magnetic fields are applied in the x, y and z directions to alter the net magnetic field strength at different locations, and hence, the Larmor frequency. By applying gradient fields at different times in the pulse sequence, in combination with T1- and T2-weighting, different contrasts can be achieved in the resulting image that allow differentiation of, for example, grey matter, WM and cerebrospinal fluid (CSF). More advanced imaging techniques allow, not only contrast between tissue

types, but also characterisation of other tissue properties, the most relevant of which to this thesis are white matter fibre structure and blood oxygenation level.

Blood Oxygenation Level Dependent (BOLD) Contrast

The magnetic properties of the haemoglobin molecule vary depending on whether or not it is bound to oxygen [124]: deoxygenated haemoglobin is paramagnetic while oxygenated haemoglobin is diamagnetic. Because the surrounding magnetic field is distorted by paramagnetic materials, the precession frequency of nearby protons will vary, and their transverse magnetisation will decay more rapidly (shortening T2*) following a radio-frequency pulse. Deoxygenated blood therefore decreases the measured T2* signal. It was noticed that this discrepancy could be used to produce an MRI contrast which was dependent on blood oxygenation [125]. During neural stimulation or activity, the relative proportions of oxygenated and deoxygenated blood vary depending on metabolism and blood flow. Therefore the T2* signal varies according to the temporal dynamics of consumption and supply (the haemodynamic response; figure 8).

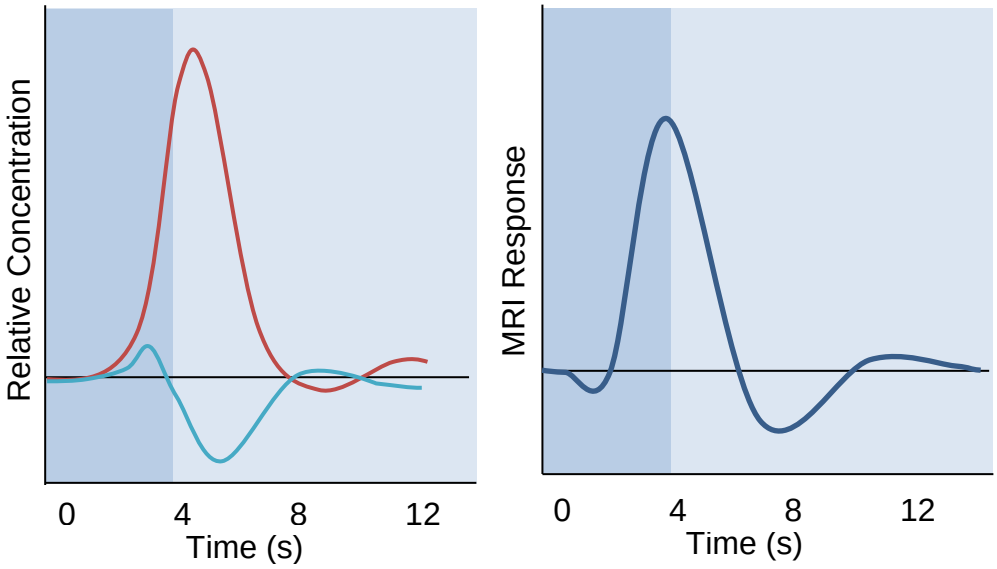


Figure 8. On the left, the graph shows the canonical relative concentrations of oxy- (red) and deoxy-haemoglobin (blue) following a 4-second neural stimulation beginning at 0s (the darker blue shaded area). The graph on the right shows the BOLD response, which is the theoretical signal observed in an MRI pulse sequence sensitive to T2* during the same stimulation. The response demonstrates the balance between supply and demand for oxygenated blood: the initial rise in deoxygenated blood due to increased metabolism is followed by a dip as oxygenated blood flushes the capillaries which supply the neurons and returns to baseline after approximately 10 seconds. Note that the MRI response is defined by the inverse of the trace for deoxygenated blood, since deoxyhaemoglobin attenuates the signal. Thus the observed MRI signal peaks not because the oxygenated blood itself increases it, but because the oxygenated blood displaces the deoxygenated blood which had been attenuating it.

The development of echo-planar imaging, among other advances in MRI hardware and pulse sequence technology, allowed volumes to be acquired much more rapidly [126]. This led to the first functional MRI (fMRI) studies of cortical activation [127, 128], in which images were acquired with a repetition time (TR) of 2 seconds to study the change in intensity in the visual cortex during stimulation. Today, fMRI is one of the central tools in psychology and neuroscience, and is applied widely in cognitive neuroscience, clinical

psychology and psychiatry, surgical planning, monitoring treatment outcomes or as a biomarker in clinical studies [129].

Diffusion-Weighted Contrast

Diffusion is the motion of molecules due to thermodynamic effects. In isotropic media, diffusion is Brownian: random and erratic, with a net motion vector of zero, $\vec{0}$. In solid or anisotropic media this motion is restricted and diffusion occurs primarily along an axis (figure 9). Diffusion-weighted MRI contrast is based on the abundance of water molecules in the body and on the inhomogeneity of the magnetic field. The MRI signal is attenuated by the loss of phase coherence (T2 effects) as the water molecules undergo varying magnetic fields.

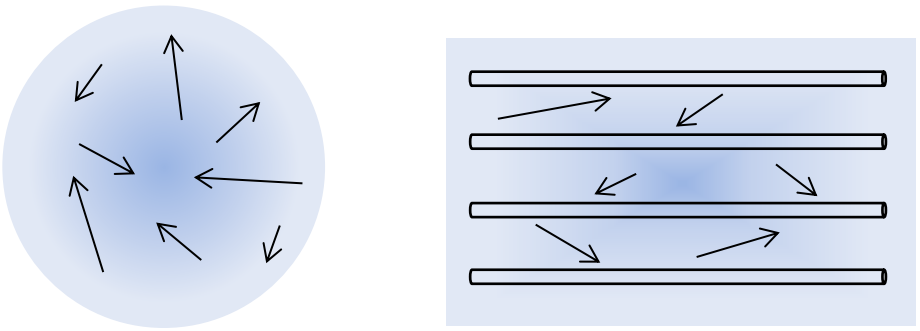


Figure 9. Isotropic (left) and anisotropic (right) diffusion. The arrows within each shape represent the motion vectors of random walk paths of water molecules. In the isotropic medium, e.g., a uniform gas or liquid, the molecules are free to move equally in any direction,. In anisotropic diffusion, e.g. close parallel fibres such as in white matter, the freedom of the molecules is greater in the axis parallel to the fibres than in the one perpendicular to them.In both cases, the net motion is zero.

Magnetic gradients can be applied to further attenuate the signal along a chosen direction. The attenuation effect due to diffusion weighting, A, is given by:

$$A = e^{-bD}$$

Here, b, known as the b-factor, is the amount of diffusion weighting applied within a pulse sequence. It depends on the strength and duration of the diffusion weighting gradient, and the gyromagnetic ratio of water. D is the apparent diffusion coefficient, which is the quantification of diffusion, assuming isotropy. Using this information, series of volumes can be acquired with weightings for different diffusion directions, from which various measures of microstructure can be derived. In the brain, diffusion-weighted contrast is typically used to investigate the structure of white matter fibres, which, because they tend to run in parallel bundles, attenuate the diffusion-weighted signal.

3.2. Imaging MS in Clinical Practice

Recent guidelines reached by expert consensus from the MAGNIMS group gave recommendations for a standardised brain MRI protocol for diagnosis of MS [130]. The suggested protocol is:

A. Axial proton density and/or T2-FLAIR/T2-weighted
B. Sagittal 2D or 3D T2-FLAIR
C. 2D or 3D contrast-enhanced T1-weighted

And optionally:

D. Unenhanced 2D or high-resolution isotropic 3D T1-weighted

E. 2D and/or 3D dual inversion recovery

F. Axial diffusion-weighted imaging

T2-weighted sequences are highly sensitive for detecting MS lesions and are recommended to be acquired in at least two planes. The sagittal T2 is acquired to allow better identification of lesions in the corpus callosum. The gadolinium contrast enhanced T1-weighted image is suggested because it allows differentiation between acute and chronic lesions, which is useful for establishing dissemination in time. For follow-up examinations, the same protocol is recommended, minus step B. The optional acquisitions are listed which could provide further useful information but may not be suitable depending on the situation.

3.3. Imaging MS in Research

A number of non-conventional MRI techniques have been applied in recent research of the MS brain. These imaging techniques may reveal important new mechanistic information about the disease process. Particularly of relevance for this thesis are functional MRI (fMRI) and diffusion MRI.

When fMRI data are acquired while a subject is at rest but awake, low frequencies (<0.1Hz) in the observed signal are known to have a physiologic origin [131]. Temporal correlations between fMRI timecourses of brain regions therefore reflect the degree of functional connectivity between them [131]. One specific group of correlated brain regions, known as the default-mode network (DMN), routinely deactivates during task performance [132] and is thought to reflect intrinsic baseline neuronal activity. These principles are useful for the study of brain function in MS because the multiple white matter

lesions determine functional disconnections between areas of the CNS. Hence most research on functional connectivity in MS has focussed on resting-state connections in the brain's intrinsic functional architecture and how they relate to clinical progression and cognitive decline [133]. Applications of fMRI for the study of brain networks in MS are reviewed later in section 4.3.

Diffusion-weighted MRI data can be used to estimate a diffusion tensor which describes the direction and magnitude of diffusion in a particular voxel [134]. When a tensor is computed for every voxel in the brain, fiber tract trajectories can be traced and used to study neural pathways *in vivo* [135]. In MS, one example of an application of this technology is the demonstration of increased diffusivity and lesion volume in the cortico-spinal tracts of patients with motor impairment [136]. It has been used to identify white matter tracts which intersect T2 lesions and are at risk of degeneration [137]. A second use for diffusion MRI is to derive summary indices of white matter microstructure; for example, fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD) and radial diffusivity (RD) [138]. These indices are defined by formulae using the three main eigenvalues of the tensor, λ_1, λ_2 and λ_3. FA is the relative difference between the largest eigenvalue and the two other eigenvalues, which describes the extent of anisotropy. MD is the average diffusion in all three eigenvalues. AD describes the diffusion in the longest eigenvalue and RD describes the average of the two shortest eigenvalues.

$$FA = \sqrt{\frac{1}{2}} \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_1 - \lambda_3)^2 + (\lambda_2 - \lambda_3)^2}}{\sqrt{(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}}$$

$$MD = \frac{(\lambda_1 + \lambda_2 + \lambda_3)}{3}$$

$$AD = \lambda_1$$

$$RD = \frac{(\lambda_2 + \lambda_3)}{2}$$

Magnetic resonance spectroscopy using hydrogen is a technique allowing the *in vivo* measurement of the concentrations of neuro-metabolites. This is potentially of interest in MS research because there are metabolic changes associated with the breakdown of myelin, remyelination and inflammation; for example, N-acetyl-aspartate is decreased by neuronal damage, choline is increased with myelin breakdown and creatine increases with cell density. Early studies have shown that, in gadolinium-enhancing lesions, creatine, choline, myoinositol and glutamate concentrations are increased, while N-acetyl-aspartate is decreased. In chronic lesions, concentrations of N-acetyl-aspartate are greatly reduced, myoinositol is increased and glutamate is unchanged [139, 140]. Another technique, magnetisation transfer (MT) imaging, is used to measure the transfer of energy from free water protons to bound water and macromolecules and can be used to accurately measure tissue damage in MS lesions [141].

4. The Graph-Theoretic Model of Brain Networks

All of nature is intricately interconnected. Many natural systems and processes can be described as groups of interconnected entities at multiple scales: examples of connectivity in nature range from the quantum, molecular and cellular at the lowest level, through to social, computer and economic interactions at the highest. Our ability to understand and model these systems is crucial to our interpretation of the physical world. However, the traditional Newtonian scientific method based on analysis of isolated parts under controlled conditions may not be adequate to describe all such complex systems as occur in nature; for example, the principle of reductionism is being challenged by theories of quantum mechanics, relativity and chaos in physics [142]. Some phenomena, i.e. those which are emergent, may only be explained holistically [143].

Complexity science is an alternative methodology, a scientifically-based holism, which is able to describe systems which are neither completely ordered nor completely disordered and which possess emergent properties. Biological networks fall into this category [144]. Analysis of networks is one of the principal methods of complexity science, and is based on the tenet that the whole network is greater than the sum of its parts. The emergent properties of a network are the ways in which it behaves and descriptors of its high-level structure. Many biological systems lend themselves to this mode of description; in particular, those which consist of distinct, communicating parts.

This chapter describes the basic algebra of graphs, the principal tool in the study of networks, and how they have been applied toward that end. It then describes the fundamental concepts of network science before finally updating the reader on the present state of knowledge on brain networks, specifically in MS and how they relate to impaired cognition and fatigue.

4.1. Graph Theory

A graph is a mathematical construct used to model the relationships within a group of objects. A network is an empirical interconnected system of discrete objects. Whereas a graph is a theoretic, abstract and purely conceptual structure, networks exist in the real world. Graphs are useful in modelling networks. Although often used interchangeably in the existing literature, the above definitions are used throughout this thesis. For graph theory definitions and terminology, I refer to the textbook of Chartrand and Zhang [145].

Building upon Gottfried Leibniz's *geometria situs* (geometry of position), the conception of graph theory is most commonly attributed to Leonhard Euler; specifically to a single paper, in which he famously showed that it was impossible to plan a path around the town of Königsberg which crossed each of the seven bridges only once (figure 10) and how such problems are solved in the general case [146]. For a field stemming from an ostensibly simple problem, graph theory is applied widely today throughout science and society.

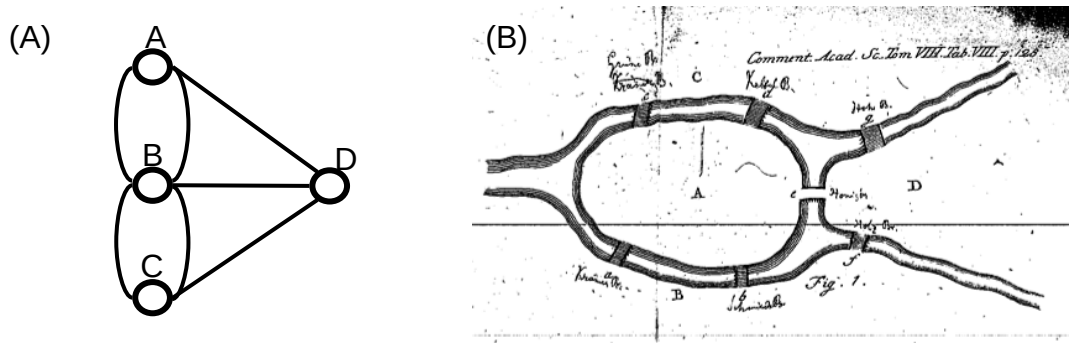


Figure 10. By convention, graphs are pictured by drawing a dot or circle for each vertex and a line between two vertices for each edge. Subfigure (A) shows the graph corresponding to the bridges of Königsberg problem reproduced from Leonhard Euler's 1736 thesis (B).

Formally, a graph G is an ordered pair of the sets V and E, where V is a finite nonempty set describing the verticies and E is a set comprising two-element subsets of V which describe the edges. For the graph in figure 10A, we write:

$$G = \{V, E\}$$

$$V(G) = \{A, B, C, D\}$$

$$E(G) = \big\{\{A, B\}, \{A, B\}, \{B, C\}, \{B, C\}, \{A, D\}, \{B, D\}, \{C, D\}\big\}$$

The number of vertices in a graph is known as its order and is denoted $|V|$. The number of edges in a graph is known as its size and is denoted $|E|$. A vertex v is said to be *incident* with edge e if $v \in e$. The degree of a vertex describes the number of edges incident upon it and is denoted $d(v)$. For example, the degrees of vertices A, B, C and D in figure 10A are 3, 5, 3 and 3 respectively, with $|V| = 4$ and $|E| = 7$.

It is often useful to encode a graph in the form of a matrix, called an *adjacency matrix*. The adjacency matrix A of a graph always has size $|V|^2$, and each element a_{ij} represents the number of edges connecting the vertices

i and j. Elements in an adjacency matrix can also represent a weight, or strength of connectivity. For graphs where connections are not directional, the adjacency matrix is always symmetric ($a_{ij} = a_{ji}$). For the graph in figure 10A, we can write:

$$A(G) = \begin{bmatrix} 0 & 2 & 0 & 1 \\ 2 & 0 & 2 & 1 \\ 0 & 2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$$

4.2. Networks

Network science is the study of pattern and structure in networks, in which graph theory is the primary tool. Network science first became established as a discipline in the 1930s when graph theory was applied to relationship modelling in sociology [147].

The first important topic in the study of networks is the modelling of random graphs. There are a multitude of ways to form a random graph and each method produces graphs with different properties. One method is the probabilistic model of graph creation by Erdős and Rényi [148], in which each edge is formed if a random number from a uniform distribution between 0 and 1 is less than a constant *p*. This model has the characteristic of a binomial degree distribution because there is no bias given to particular nodes. A second is Watts and Strogatz's procedure for generating graphs with the "small-world" (SW) characteristic (figure 11) [149]. The third well-known random network model is Barabási and Albert's [150] preferential attachment model. Under this model, new edges are more likely to connect with nodes of higher degree. This produces graphs with a degree distribution following a

power law (i.e. the degree distribution is a straight line on a log-log plot), with few high-degree hubs and many low-degree nodes.

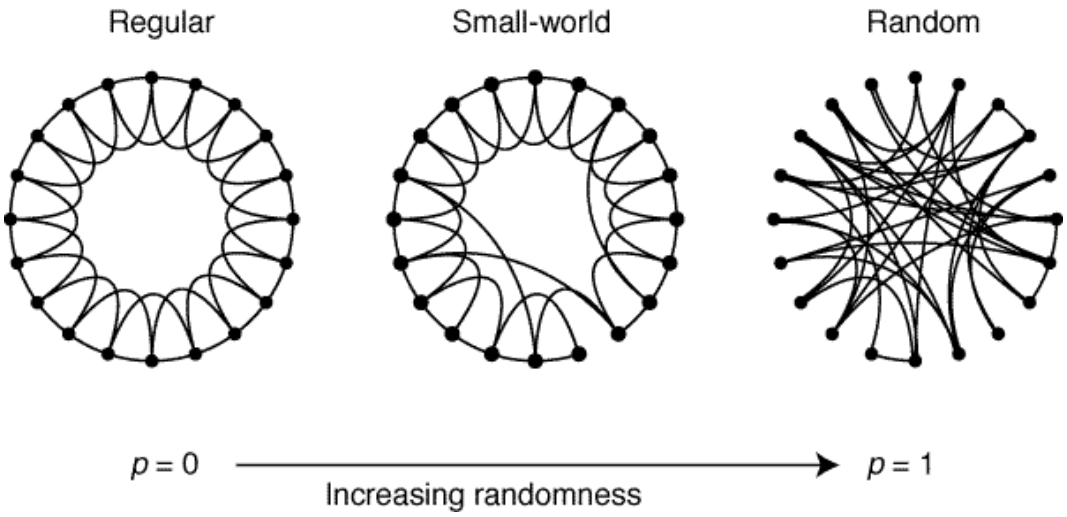


Figure 11. Watts and Strogatz's procedure for generating graphs with intermediate regularity and SW structure. Starting from a regular graph in which each node is connected to k neighbours, edges are rewired randomly with a probability *p* until each edge has been considered. The result is a network model which can be tuned through the middle ground of regular and random. Image from Watts and Strogatz [149].

From a graph, it is possible to calculate summary metrics which describe its topology. Here I will list the important and most relevant metrics:

- Density. The density of a graph is the ratio of the number of existing edges to the number of possible edges. Density is often used as a way to "threshold" a graph, removing weak connections which are thought to be unimportant, rather than applying an absolute threshold.
- Characteristic path length (CPL). Given two nodes, i and j, the shortest path between them is the path with the fewest steps. The characteristic

path length L is the average of all shortest paths across a whole network.

- Clustering Coefficient (CC). Clustering coefficient of a single node is the proportion of pairs of its neighbours (connected nodes) that are connected to each other. It is given by:

$$C_i = \frac{2N}{k_i(k_i - 1)}$$

where k_i is the degree of node i (the number of edges connected to it) and N is the number of connections existing between neighbouring nodes. The global clustering coefficient is the average of clustering coefficients across all nodes in the network.

- Small-Worldness. Small-worldness is the property of being highly clustered and being able to reach any node in a small number of steps from any other. Small-worldness is quantified with the equation:

$$\omega = \frac{\left(\frac{C}{C_{random}}\right)}{\left(\frac{L}{L_{random}}\right)}$$

Where C_{random} and L_{random} are the mean clustering coefficient and characteristic path length of Erdős-Rényi random graphs with the same size and degree sequence.

- Modularity (MOD). Graphs can be divided, for example with Newman's spectral optimisation method [151], into subsets of nodes that connect more within themselves than they do to the rest of the graph (called modules). The strength of this division can then be used as a summary measure of modularity, defined as the fraction of edges in the selected

modules minus the fraction of edges in the modules if the edges in the network were assigned randomly.

- Efficiency. The efficiency of a graph quantifies how efficiently it can pass information around and can be measured at either a per-node or whole-graph scale. The local efficiency of a graph is given by:

$$E_{loc}(G) = \frac{1}{n}\sum_{i \in G} E(G_i)$$

Where G_i is a subgraph of the neighbours of node *i* only, excluding node *i*. The global efficiency (GE) is the ratio of the average efficiency of the graph to the efficiency of a fully connected graph of the same size, where the average efficiency is given by:

$$E(G) = \frac{2}{n(n-1)}\sum_{i<j \in G} \frac{1}{d(i,j)}$$

And $d(i,j)$ is the shortest path between nodes *i* and *j*.

- Betweenness Centrality. The betweenness centrality of a node is a measure of its relative importance within a graph. It represents the number of shortest paths which traverse a given node. The betweenness centrality of a single node is:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Where σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of shortest paths through node v.

4.3. Brain Networks Relevant to Cognition and Fatigue in MS

The human brain is the quintessentially complex biological system. Given the crucial role of white matter tracts in the flow of information around the brain,

their vulnerability to focal and diffuse damage in MS and the current prevalent understanding of brain function as being dependent on the intercommunication of networks of grey matter regions, brain networks in people with MS have been a prime target for investigation. Most research in this area has focussed on the questions of “which type of changes to functional connectivity or functional activation are related to disease status and to cognitive functioning?”, and “which changes reflect adaptive or maladaptive functional reorganisation?”

The earliest studies of brain function in response to cognitive task performance in MS noticed that (a) brain regions relevant to the task were hyper-activated and (b) additional brain regions not usually thought to be relevant were recruited during task performance. These changes were more pronounced in patients with greater structural damage, but, crucially, also more pronounced in patients with better cognitive test scores. Likewise, the patients with poorer cognitive performance had decreased cortical activation in the relevant areas. This suggested that increased activation and recruitment of cortical areas may have been a favourable process. In the first of these task-activation studies, Staffen, Mair [152] had 21 subjects with early MS perform a visual analogue of the PASAT during an fMRI acquisition. They found no significant difference in MSFC or PASAT scores between patients and matched controls, but that the pattern of cortical activation differed between groups, with patients’ peak activation being in the right frontal cortex and left angular gyrus, while control subjects’ main area of activation was the right anterior cingulate. In a similar design, Audoin, Ibarrola [153] studied activation patterns during PASAT performance using fMRI in patients with

CIS, finding greater frontopolar activation in patients, with a stronger effect in those performing well on the task and no significant difference in task performance between patients and controls. A third study [154] confirmed this effect in RRMS patients, with greater frontal activation in patients with better cognitive performance. Subsequent studies have made similar findings in tasks focussing on other aspects of cognition. During an encoding and retrieval memory task designed specifically to activate the hippocampus, cognitively preserved patients had increased parahippocampal, frontal and cingulate activation compared to cognitively impaired patients [155]. fMRI activation during performance of the n-back task was greater in the right dorsolateral prefrontal cortex in cognitively preserved patients [156]. In addition to task activation, studies of functional connectivity at rest in functional networks, including the DMN, appeared to concur. Roosendaal, Schoonheim [157] showed that patients with CIS tended to have increased synchronisation in resting state networks compared to both RRMS and healthy controls. Further, patients with progressive MS had reduced activity in brain networks identified using independent component analysis [158]. Together, these observations led reviewers to the conclusion that increased activation and connectivity within the brains of people with MS was a compensatory adaptation of brain connectivity in order to maintain normal cognitive functioning [159].

However, a number of more recent contradictory results from studies of brain connectivity in MS bring this conclusion into doubt by showing the opposite: that cognitive impairment is either related to increased functional connectivity or to both increases and decreases in specific regions. Several studies have

focussed on connectivity changes in the DMN, finding both increased and decreased connectivity. For example, in multiple studies in which patients were classified as either cognitively preserved or impaired, the patients with cognitive impairment tended to show decreased resting-state network integration in the DMN [160-163]. On the other hand, studies have shown negative correlations between measures of function and increased network connectivity in the DMN [164, 165]. For the DMN to negatively correlate with cognition does not fit with current models; however, one of these studies only found such a relationship for structural connectivity [165] and the other was conducted in a very small sample and may be an anomaly [164]. Others have found increased connectivity of key regions of the DMN during rest; for example, in relation to processing speed, specifically in the ventromedial prefrontal cortex and frontal pole [166] and at rest bilaterally in inferior parietal cortex, posterior cingulate and medial prefrontal cortex [167]. In paediatric MS, reduced connectivity in cognitively impaired patients was found specifically in the precuneus [168]. One study showed areas of both positive (cerebellum, left middle temporal gyrus, left occipital pole and left angular gyrus) and negative (right pars opercularis, right inferior frontal gyrus and right caudate) correlation between PASAT performance and resting-state functional connectivity [169]. Two studies have also highlighted connectivity changes in the dorsolateral prefrontal cortex (DLPFC) relevant to cognition. In cognitively preserved patients, performance on the n-back task was related to increased connectivity of the DLPFC to medial frontal cortex [170] while connectivity between the same regions was decreased in another study during a “go/no-go” task [171]. Several studies have also shown various

changes in frontal and dorsolateral prefrontal connectivity during PASAT performance [172-174].

With both sides of this evidence in balance, the taxonomy of adaptive vs maladaptive changes in MS employed throughout the cited research may, therefore, not be sufficient to describe the complex simultaneous processes of neuronal damage, remyelination, network dysfunction and plasticity. Further, the usage of the term “maladaptive” plasticity itself may be problematic when applied in MS, in that a definition is just beginning to be established, even in neuroscience generally [175]. This research has, however, gone some way toward localising the changes in functional activation and remapping which are relevant to cognitive impairment. Speculatively, the key to understanding these apparently conflicting reports may be to look at the stage of the disease in the studied cohorts: it appears that, early in the disease course, the capacity of the brain to reorganise and cope with damage is at its highest. In studies of brain changes in early MS, a predominant shift in activation and connectivity toward the frontal cortex is observed during cognitive task performance. At the onset of clinically-relevant cognitive impairment, brain networks have exhausted their capacity to adapt to damage and network efficiency therefore declines more rapidly. The point at which this occurs depends on the extent of structural damage and the redundancy or resilience of the networks underlying cognition. This putative timeline is shown in figure 12. The idea of a protective functional reserve against network disconnection in MS is new and relatively unexplored but potentially important for the collective understanding of cognitive impairment [170, 176].

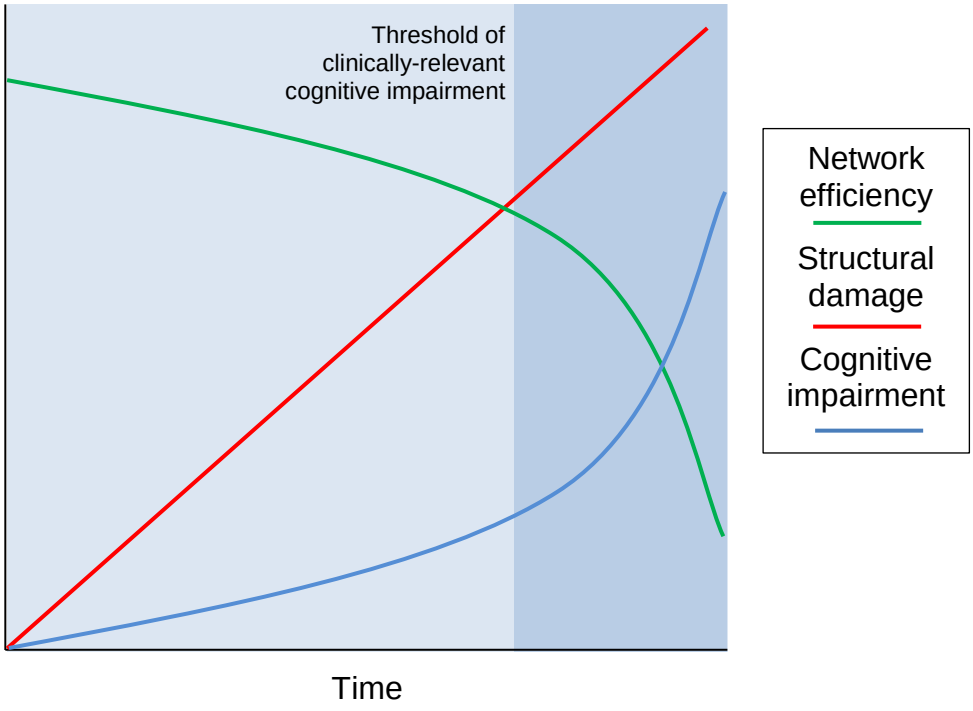


Figure 12. After [34], the graph shows a hypothesis of the effect of accumulating structural damage on brain network efficiency and cognitive impairment. The shaded area represents the onset of clinically-relevant cognitive impairment.

The graph theoretic network model has been proposed as an alternative, more holistic model to the task-activation and region-specific connectivity models reviewed above. Over the previous attempts to investigate brain functional changes in MS, the graph model has the advantage of being able to look beyond local changes and to instead see how the connectivity of the entire network is impacted. A total of 17 known articles have reported on the application of a graph-theoretic model to MRI data in people with MS. These studies have used a range of data types for the construction of graphs, including cortical thickness, tractography streamline counts and functional connectivity based on fMRI, magnetoencephalography (MEG) or electroencephalography (EEG) data. Samples of people with CIS and all

subtypes of MS, with varying disease durations and levels of impairment have been studied with the technique.

In the first such publication, He and colleagues [177] reported that the small-world property of brain networks in people with MS was disrupted (reduced) and that the reduction correlated positively with lesion load. The measure of connectivity used was the Pearson correlation coefficient between the cortical thickness of pairs of regions across all subjects, yielding one graph per group (the cohort of 330 patients was split into 6 equal groups based on lesion load). The first step in the study was to show, as is convention, that the constructed networks had the typical features of a small-world network, falling somewhere between regular and random networks for measures of local and global efficiency over the range of graph thresholds (figure 13 A&B). The main outcomes of the study were the correlations between graph metrics and lesion load. This was done by testing the correlation of the mean lesion load of each of the 6 groups with the graph metric scores (figure 13). This was the first and only paper using covariance of cortical thickness measurements as a connectivity measure and, while the results led to many further investigations, the biological relevance of this measure is not clear and the findings are therefore difficult to interpret.

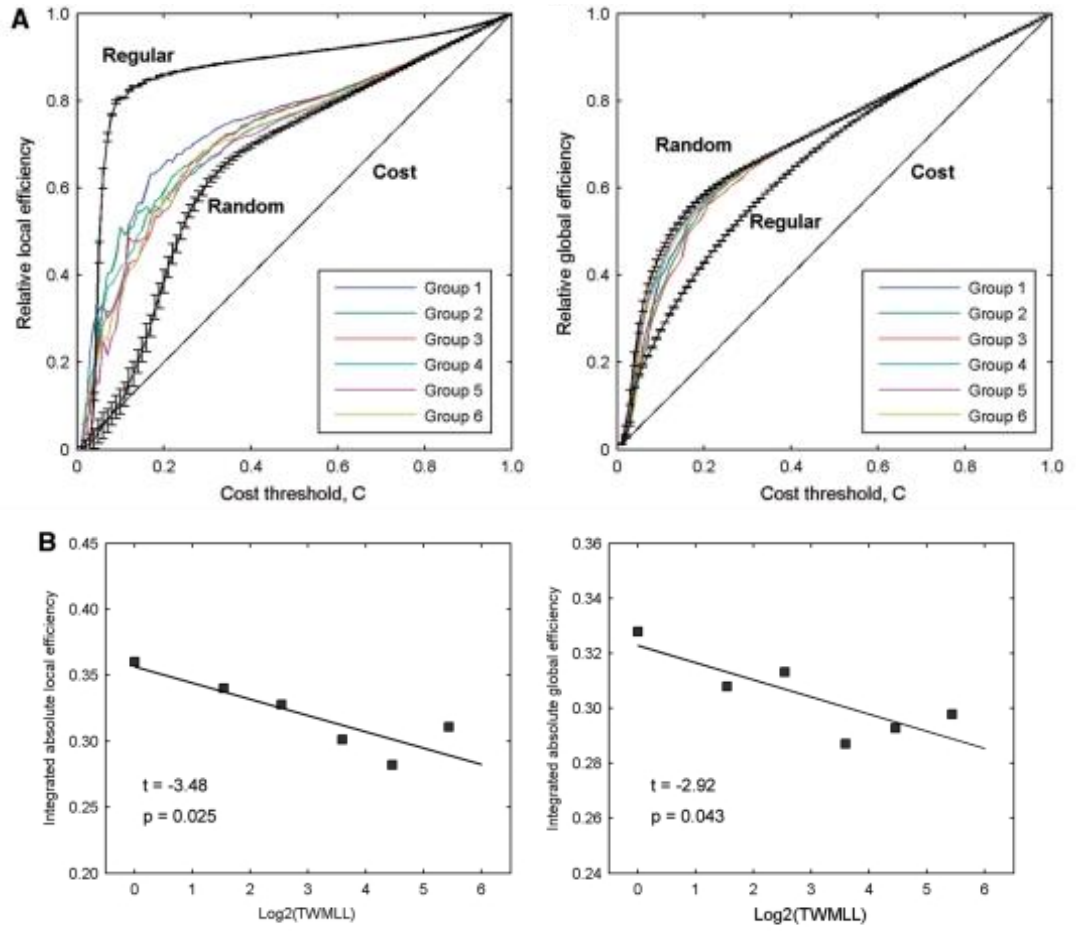


Figure 13. Taken from [177], the figure shows (A) the authors' demonstration that the studies human networks lay between constructed random and regular networks in terms of local and global efficiency and (B) the main outcome of the study showing significant negative correlations between network efficiency and white matter lesion load.

A number of studies then investigated the organisation of brain grey matter networks, predominantly in RR and SP MS, using graphs based on diffusion-tensor imaging with tractography streamline counts as the connectivity measure. However, the results of these studies are conflicting, even where similar methods are applied and where the studied cohorts are similar in disability measured by the Expanded Disability Status Scale (EDSS), age and subtype. For example, the study of Kocevar, Stamile [178] found significantly higher transitivity and characteristic path length in patients, while the study of

Llufriu, Martinez-Heras [179] found that the same metrics were significantly lower in patients. Other studies found increased modularity localised to the cerebellum, cingulum and temporo-parietal regions [180], reduced communicability in frontal, parahippocampal, motor and occipital regions [181], and decreased global efficiency in regions subserving motor, visual and language functions [182].

A number of groups have also investigated brain network organisation in MS using graphs based on functional connectivity measures, most of which are based on resting-state fMRI. The most reproduced finding, as shown in 4 independent studies [183-186], was a reduced global network efficiency in MS patients compared to controls. Moreover, patients with CIS tended to have intermediate efficiency scores falling between the scores of healthy control and MS groups [183, 186]. Interestingly, low network efficiency was also found in three of the studies investigating structural networks in MS [178, 179, 186]. Other studies showed increased characteristic path length [184, 186], increased modularity [187], and regional variations in measures of centrality [188]. Another subset of studies used magneto- and electro-encephalography to study brain networks in people with MS. These studies have identified group differences in the centrality of network hubs [189], increased path length and clustering in patients in the lower alpha band [190], lower global integration in alpha and beta bands in patients [191], loss of hierarchy in alpha2 band networks associated with poorer cognition [191] and that networks in the theta band tend to be more regular but more random in the alpha band [192].

Most of the reviewed articles applying graph methods in data from people with MS also aimed to test the relationships between brain network topology and cognitive impairment. The most common measure of cognitive ability among these studies was the PASAT. One study of fMRI in patients with CIS found no significant relationships between network efficiency and PASAT [183]. Others found that better PASAT performance was related to increased nodal strength [179], local and global efficiency [193], increased clustering [193], lower mean degree, global efficiency and hierarchy scores [194], higher path length [194], increased modularity at medium densities [187], increased clustering in the theta and delta ranges in frontal areas [195] and higher clustering in the lower alpha band [190]. Some studies formed summary cognitive scores based on multiple tests and found relationships between cognition and eigenvector centrality in the ventral stream [196], greater nodal centrality in temporal regions [189] and network hierarchy in the alpha2 band [191]. Another study reported significant correlations between network efficiency and visuo-spatial memory in male patients only [185]. One study reported no relationship between SDMT scores and any measure of network organisation [179].

Together, these studies have gone some way toward showing that brain graph metrics may be valid measures of various types of functional impairment in MS at the group level. There is some evidence of a consensus that brain networks in people with MS tend to less efficient than in controls, and that graph measures of segregation are increased and measures of integration decreased. There is also some evidence of a relationship to cognitive performance. Notably missing from this literature are analyses of

reliability across time, subjects and scanners, and responsiveness to changes in cognitive state.

Part II: Research Studies

5. The Relevance of White Matter Degradation to Impaired Function in Multiple Sclerosis [1]

5.1. Introduction

Magnetic resonance imaging is used extensively for understanding the mechanisms behind disability in MS [197, 198]. Investigation of the cerebral white matter is especially pertinent because of the propensity for MS to affect the white matter. Previous studies that mapped white matter lesions have identified associations between lesion distribution and disability, but do not account for the widespread diffuse WM damage present in NAWM [199-204]. Diffusion tensor imaging (DTI) allows quantification of FA, which is sensitive to this damage [202, 205]. FA is a marker of ultrastructural WM integrity and has been used for more than a decade to study WM lesions and damage to NAMW [202, 205-209].

Since its first appearance in 2006 [210], tract-based spatial statistics (TBSS) has been used for multisubject spatial analysis of DTI data to investigate neurologic and cognitive correlates of WM tract degradation in MS [210-227]. These studies demonstrate differences in patterns of tract-based functionally

[1] This work has been previously published: 36. Welton, T., et al., *Functionally Relevant White Matter Degradation in Multiple Sclerosis: A Tract-based Spatial Meta-Analysis.* Radiology, 2015. **275**(1): p. 89-96. My role in the publication was lead author, responsible for design, systematic review, data collection, analysis, manuscript preparation and submission. Other contributors to the publication were Dr Rob Dineen (concept and design), Mr Daniel Kent (repeat systematic review), Prof Dorothee Auer and Prof Cris Constantinescu. All authors reviewed and commented on the final manuscript.

relevant FA change; for example, correlations between tract FA and measures of cognition or physical disability. The studies generally support the idea that disconnection of cortical or subcortical grey matter by diffuse or focal damage to interconnecting WM tracts may be an important contributor to physical or cognitive disability in MS [214]. However, general interpretation of these studies is limited by heterogeneity of methods and demographics, including cohort age and level of disability, field strength of the magnet and the number of diffusion gradient directions used.

From these published studies, some common findings can be identified. First, reduced FA and increased MD in the corpus callosum and pyramidal tracts was reportedly associated with greater physical disability [210, 215, 217-219]. Second, an association between poorer cognitive performance and lower FA in the corpus callosum, posterior thalamic radiation and posterior cingulum was reported [35, 216, 217, 219, 220, 226, 227].

Because the sensitivity of DTI varies depending on the direction and density of nerve fibres [228], spatial meta-analysis of these studies using unthresholded t-statistic maps could increase statistical power and allow identification of potentially important regions of interest which would otherwise go unnoticed, such as in smaller tracts. I therefore aimed to identify statistical consensus between published studies for the distribution and functional relevance of tract WM degradation in MS.

This chapter provides evidence toward research question 1: “Are cognitive impairment and fatigue in MS the result of disruption to macroscopic brain networks?”

5.2. Methods

Literature Search

First, the same literature search was performed independently by two different researchers. Performing two parallel searches allowed us to eventually compare results and produce a consensus on the final selection with fewer errors. The MEDLINE, Web of Knowledge and Google Scholar databases were searched using the string:

> “multiple sclerosis” AND ((“TBSS” OR “tract-based spatial statistics”) OR (“DTI” OR “diffusion tensor”))

The title, abstract and keywords fields were searched. These databases were chosen for the greatest coverage [229]. All languages were included and results were filtered to include only articles published during or after 2006, which is the year of publication of the original TBSS methods paper by Smith, Jenkinson [210]. These searches were performed in January, 2014. Abstracts and, where necessary, full texts of the resulting articles were screened to identify only those that performed TBSS analysis of DTI data. A second stage of screening identified articles that included one or more of the following voxelwise analyses:

a) Group comparison of FA in MS to FA in healthy controls.
b) Correlation of FA in MS to scores on the EDSS.
c) Correlation of FA in MS to any general or summary measure of cognition, including the PASAT.

Articles were excluded if they met any of the following criteria:

- Focussing on individual regions of interest, instead of whole-brain analyses.
- Comparing whole-brain summary DTI measures, like mean FA, with test scores instead of performing voxelwise tests.
- Reporting results from the same cohort as another article. In cases where this was not clear, the corresponding authors of the articles were contacted to confirm.

We searched the reference lists of the remaining selection for any other relevant articles and finally, between the two researchers performing the search, came to a consensus on a final selection by discussion. From these articles, I tabulated the following details:

- Numbers of people with MS and controls, which were required for meta-analysis.
- Numbers of each MS subtype in the MS group for reader clarity on the generalisability of the results.
- Country of origin of the sample, to show generalisability across a wide geographical distribution.
- Sample mean ages for testing possible moderating effects by meta-regression.
- Number of diffusion gradient directions in the DTI sequence for meta-regression.
- Which of the three voxelwise comparisons were made, so the relevant data could be gathered and used in the correct meta-analyses.
- Mean EDSS score of the MS group for meta-regression.

- Magnet strength for meta-regression.
- Mean lesion volume for meta-regression.

The parameters for meta-regression were chosen based on their availability in the published literature. Other measures could have been used, such as b-value, number of averages, voxel size or SNR, but were not always reported.

The corresponding author of each of the articles was contacted by email to request the unthresholded *t-statistic* maps from the output of their TBSS analysis and any data in the above list which was missing from the published article. Where possible, I gathered both the “tstat1” and “tstat2” image outputs from each comparison, which represent group1>group2 and group2>group1, respectively, and confirmed which image should be used in the meta-analysis, as some investigators may have ordered their groups differently. All images were inspected by eye to ensure that the correct output had been submitted, as some of the authors had renamed the files or submitted the wrong files initially.

Quality Assessment

I assessed the methodologic quality of each article against a set of nine weighted a-priori criteria. The criteria (table 2) were based on those of Kmet, Lee [230], but adapted to meet the needs of this review. The weights of the criteria were designed to give importance to image quality, rather than quality of reporting. Articles scoring lower than the predefined threshold of 50%, corresponding to fewer than 9 of the 17 available points, were excluded from further analysis.

Table 2. Quality assessment criteria.

Criterion	Weight
Did the authors have a clear a priori hypothesis and design?	3
Did all members of the MS group have a clinically-definite diagnosis?	1
Were the treatments currently being received by members of the MS group recorded?	1
Did the authors justify their chosen FA threshold?	1
Did the DTI protocol use 20 or more diffusion directions?	3
Did the scanner used have a magnet strength of 3 Tesla or greater?	3
Did members of the MS group undergo clinical assessment at the time of participation?	1
Did all subjects receive the same intervention using the same facilities?	2
Were the data processing steps appropriate considering the hypothesis?	2

Meta-Analysis

Effect Size Signed Differential Mapping software (ES-SDM version 4.12; http://www.sdmproject.com/; now called seed-based d mapping; Radua, Mataix-Cols [231]) was used to perform the three separate spatial meta-analyses. ES-SDM allows efficient voxelwise meta-analyses through the inclusion of both peak coordinates and statistical maps. In cases where statistical maps are available, the creation of unbiased effect size and variance maps is straightforward [232]. Where peak coordinates are used, the maps are recreated by applying a 20mm full-width half maximum kernel to each peak. Voxels closer to a peak were therefore assumed to have a greater effect size.

Additional processing steps were needed in order to create effect size and variance maps [233] for TBSS studies. First, 100,000 positive and negative low-thresholded (±0.1) local maxima were retrieved from each statistical map (Appendix A) and input to the SDM peak coordinates procedure, described above, to create effect size and variance maps. This ensured that all effect size maps correctly overlapped the TBSS template.

A voxelwise mean of the study maps was created, which was weighted by the mean of the inverse of each study's variance and the inter-study heterogeneity. This meant that the results accounted for study size and for brain regions that were highly variable across studies. Voxel-based permutation tests determined significance. Based on an empirical validation of SDM by its creators [231], I used the recommended thresholds which were determined to balance sensitivity and specificity, and to approximately correspond to an equivalent corrected p-value of 0.05 ($z > 1$; cluster extent ≥ 10 voxels; uncorrected $p < 0.005$).

False positives are a recognised problem in voxelwise meta-analyses [234]. To assess the effects of false-positive results, I also ran the meta-analyses with a more stringent p-value of 0.00001. I also conducted leave-one-out sensitivity analyses to assess the robustness of the findings, in which the meta-analyses were repeated, leaving-out one study at a time.

I performed meta-regressions, which are a built-in function of the SDM software, to assess the moderating effects of age, EDSS score, magnet strength, lesion volume and number of diffusion directions. For the meta-regressions, I followed the guidelines of the SDM authors in using a p-value

threshold of 0.0005, only including clusters that were present in both the slope and in one extreme of the regressor, and ignoring clusters that were not also present in the main meta-analyses.

5.3. Results

Literature Search

Figure 14 shows the results at each stage of the literature search process. Of 127 search results, 36 articles (28.3%) were duplicates of searches in other databases. Of the 91 unique search results, 68 articles (75%) were excluded because they did not analyse diffusion data from people with MS using TBSS. Five further articles were excluded for not meeting the other inclusion criteria. The corresponding authors of the remaining 18 articles were contacted. All corresponding authors responded. For five of the studies, the images required were no longer available [212, 222-225] and one further study [213] was excluded because it used the same dataset as another [210]. I was also supplied with additional unpublished data for the FA-to-EDSS comparison by two authors [35, 227]. The final dataset comprised 495 people with MS and 253 healthy controls from 12 studies across 10 countries (table 3; figure 14) [35, 210, 211, 215-221, 226, 227].

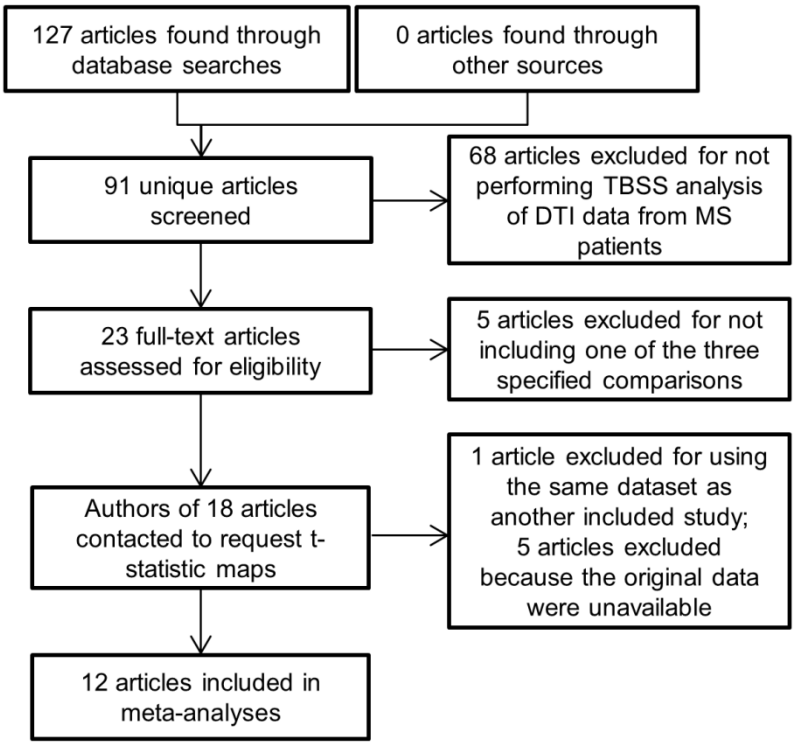


Figure 14. Flowchart summarising the literature search process.

Table 3. Studies included in the final meta-analyses.

	MS				Control			Study						
												Comparisons		
Reference	N	Age	%F	EDSS	N	Age	%F	Country	Subtypes	Scanner T	Diffusion Directions	FA-cognition	FA-physical disability	MS-controls
[211]	14	15.0	73	0.75	14	14.7	73	Germany	-	3	20			✓
[35]	37	43.5	88	3.00	25	36.4	66	UK	35 RR, 2 SP	3	15	✓	✓	✓
[215]	45	29.0	64	1.50	0	-	-	Austria	45 RR	1.5	6		✓	
[218]	41	36.8	54	2.00	41	34.6	54	China	41 RR	1.5	6		✓	✓
[219]	67	39.5	64	1.50	26	36.0	65	Spain	67 RR	3	30	✓		✓
[221]	8	37.0	63	3.50	12	40.0	58	Romania	-	1.5	25			✓
[226]	131	40.5	67	1.50	49	41.0	59	Netherlands	114 RR, 8 PP, 9 SP	3	30	✓		✓
[210]	15	43.0	53	2.50	0	-	-	UK	13 RR, 2 SP	1.5	60		✓	
[227]	37	40.9	84	2.25	20	34.0	80	USA	37 RR	3	15	✓	✓	
[220]	20	42.4	100	2.25	20	42.5	100	Canada	20 RR	1.5	55	✓		✓
[216]	55	50.2	55	4.0	30	44.5	63	Netherlands	39 RR, 16 SP	1.5	60	✓		✓
[217]	25	37.0	76	1.7	16	34.0	81	USA	25 RR	3	12	✓	✓	✓

Abbreviations: RR = relapsing-remitting; SP = secondary progressive; PP = primary progressive.



Figure 15. World map showing the geographical locations of the included samples. Each dot represents a study or group of studies where they overlap, and is approximately scaled by the size of the sample. Latitude is strongly associated with MS prevalence [235]. Latitude-wise, our sample ranged from Los Angeles (latitude ~34) to London (latitude ~52), but this is likely representative of the number of MRI studies in MS, and not the prevalence of MS due to latitude.

Quality Assessment

Table 3 shows the results of the quality assessment. The mean quality score was 76.7% ± 14.1 (standard deviation), corresponding to 13 of 17 points. All of the included articles surpassed the minimum quality threshold of 50%, or 9 points.

FA in MS and FA in Controls

A comparison of FA in the MS group to FA in controls was made in nine of the twelve included articles. The combined sample of these studies was 398 people with MS and 233 controls. The meta-analysis revealed widespread

lower FA in MS, predominantly in the corpus callosum, periventricular WM and fornices (figure 16). There was one large cluster which included all those regions (4379 voxels; $z = 7.1$; $p < 0.001$) and seven small clusters. There was one cluster of higher FA in the MS group in the right posterior internal capsule (24 voxels; $z = -1.3$; $p < 0.001$).

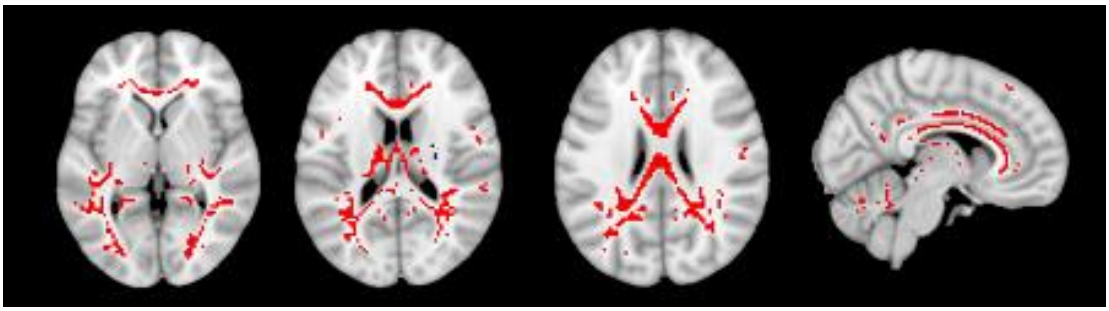


Figure 16. Comparison of FA in MS to FA in healthy controls. The images show the weighted mean across studies overlaid on the 1mm MNI 152 brain. Red voxels show areas in which lower FA for individuals with MS was significantly associated with disease status, and blue voxels, higher FA (from left to right, the slice coordinates are: z=75, z=90, z=98, x=100).

Table 4. Significant clusters for the comparison between FA and MS diagnosis.

Cluster Group Name & Sub-Clusters (a)	MNI Coordinates	SDM Z-Value (b)	P-Value (c)	Number of Voxels (d)	Cluster Breakdown (Number of Voxels) (d)
Lower FA associated with MS diagnosis (MS < controls)					
Corpus callosum body	16, -26, 32	7.139	<0.000000001	4379	Corpus callosum body (531) Corpus callosum genu (307) Corpus callosum splenium (304) R posterior thalamic radiation (159) BA 20 (138) L posterior thalamic radiation (128) L anterior corona radiata (125) BA 18 (125) L thalamus (102) R posterior corona radiata (97) BA 37 (92) R anterior corona radiata (91) R sagittal stratum (88) BA 20 (81) L fornix crus / stria terminalis (80)
Cerebellum	4, -56, -18	4.997	0.000001	160	Middle cerebellar peduncle (40) R inferior cerebellar peduncle (12) BA 37 (11) BA 18 (11)
BA 10	14, 58, 10	4.704	0.000008	48	BA 9 (16) BA 31 (11)
BA 43	56, -10, 24	4.561	0.00001	54	BA 47 (13)
BA 9	-10, 36, 46	4.415	0.00002	44	BA 31 (12) BA 9 (11)
BA 47	-46, 14, 12	4.314	0.00004	59	BA 47 (23) BA 5 (15) BA 47 (11)
R cerebellum lobule VI	8, -68, -24	4.080	0.0001	20	R cerebellum lobule VI (14)

BA 22	58, -30, 10	3.984	0.0001	20	BA 41 (10) BA 22 (10)
Higher FA associated with MS diagnosis (MS > controls)					
R posterior internal capsule	26, -16, 14	-1.322	0.0000009	24	R posterior internal capsule (21)

(a) Cluster group names assigned by SDM are illustrative and do not necessarily describe contiguous clusters.
(b) Voxel probability threshold: p = 0.005.
(c) Peak height threshold: z = 1.
(d) Cluster extent threshold: 10 voxels. Regions with fewer than 10 voxels are not included in the cluster breakdown. Cluster breakdown only includes the 15 largest regions where more than 15 exist.

FA and Physical Disability

A voxelwise correlation between FA and EDSS score in the MS groups was reported in six of the twelve included articles. The combined sample comprised 200 people with MS. More severe disability was associated with lower FA in one large cluster which included the posterior body and splenium of the corpus callosum, left fornix crus, bilateral thalami, posterior thalamic radiation and stria terminalis (323 voxels; $z = 1.7$; $p < 0.001$). There were no clusters in which lower FA was associated with lower levels of disability.

Table 5. Significant clusters for the comparison between FA and EDSS scores in MS.

Cluster Group Name [a]	MNI Coordinates	SDM Z-Value [b]	P-Value [c]	Number of Voxels [d]	Cluster Breakdown (Number of Voxels) [d]
Lower FA associated with greater disability (FA < EDSS)					
L Fornix crus / stria terminalis	-30, -22, -8	1.701	0.00003	323	Corpus callosum splenium (96) L fornix crus / stria terminalis (37) L thalamus (26) Corpus callosum body (25) R posterior thalamic radiation (18) BA 18 (12) R thalamus (11)
Lower FA associated with less disability (FA > EDSS)					
(none)					

(a) Cluster group names assigned by SDM are illustrative and do not necessarily describe contiguous clusters.
(b) Voxel probability threshold: p = 0.005.
(c) Peak height threshold: z = 1.
(d) Cluster extent threshold: 10 voxels. Regions with fewer than 10 voxels are not included in the cluster breakdown.

FA and Cognition

A voxelwise correlation between FA and cognition was reported in seven of the twelve included studies. The combined sample comprised 417 people with MS. All studies used the PASAT as a measure of cognition except the study of Schoonheim, Vigeveno [226] which used a composite measure derived from seven cognitive domains, and the study of Mazerolle, Wojtowicz [220] which used the SDMT. More impaired cognition was associated with lower FA in one large cluster which included the thalami, fornices, genu of the corpus callosum, right cingulum and right posterior thalamic radiation (1073 voxels; $z = 2.5$; $p < 0.001$). No clusters showed lower FA being associated with better cognitive performance.

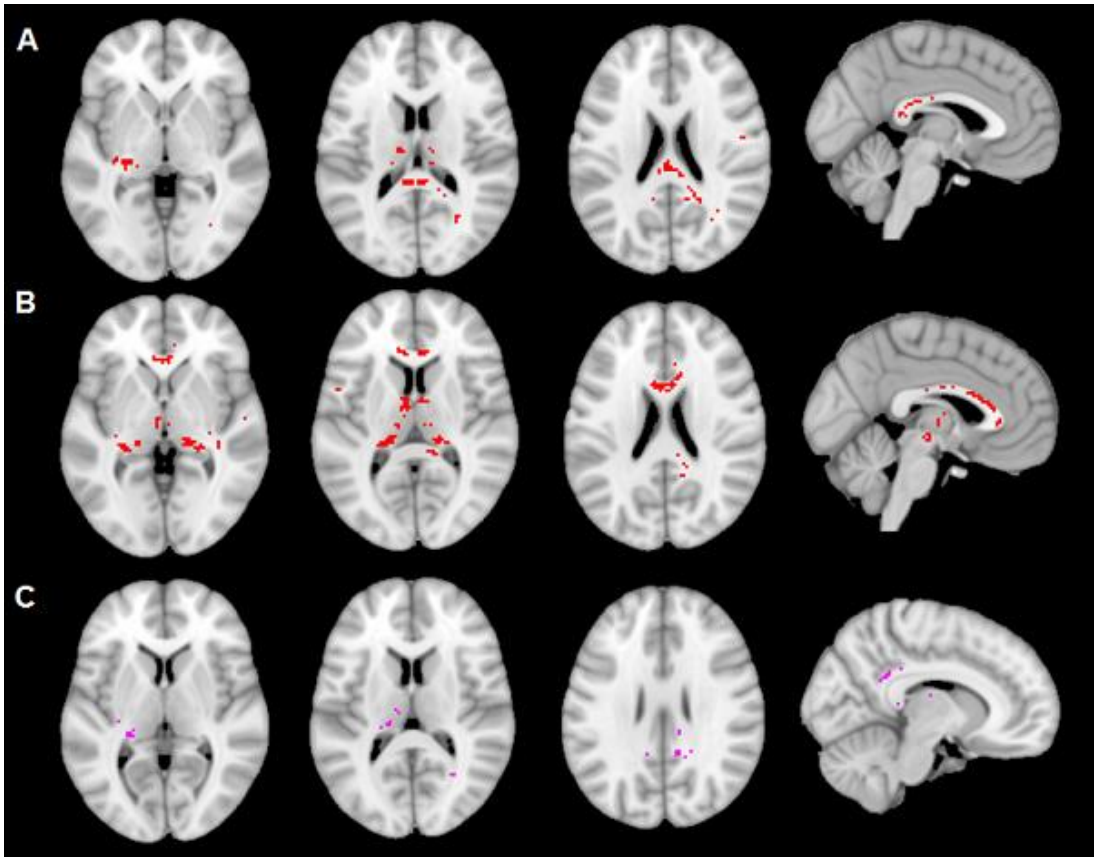


Figure 17. Significant regional correlations between FA and measures of physical disability and cognition. The images show the weighted mean across studies overlaid on the 1mm MNI 152 brain. In part A, red voxels show areas in which lower FA was significantly associated with greater physical disability (from left to right, the slice coordinates are: z=72, z=88, z=95, x=90). In part B, red voxels show areas in which lower FA was significantly associated with more impaired cognition (from left to right, the slice coordinates are: z=73, z=82, z=95, x=90). Part C shows, in pink, the significant (positive) voxels common to both physical disability and cognition correlations (from left to right, the slice coordinates are: z=78, z=86, z=100, x=102).

Table 6. Significant clusters for the comparison between FA and cognitive test scores in MS.

Cluster Group Name [(a)]	MNI Coordinates	SDM Z-Value [(b)]	P-Value [(c)]	Number of Voxels [(d)]	Cluster Breakdown (Number of Voxels) [(d)]
Lower FA associated with poorer performance (FA < test scores)					
Fornix crus / stria terminalis	-24, -34, 4	2.532	<0.000000001	980	Corpus callosum genu (211) Corpus callosum body (199) L thalamus (103) Corpus callosum splenium (75) R thalamus (68) R cingulum (36) L fornix crus / stria terminalis (31) R fornix crus / stria terminalis (19) BA 22 (14) L cingulum (13) BA 48 (12)
R posterior thalamic radiation	28, -64, 14	2.013	0.00003	19	R posterior thalamic radiation (16)
L posterior thalamic radiation	-34, -58, 14	1.998	0.00004	42	L posterior thalamic radiation (12)
BA 20	-44, -10, -30	1.900	0.0001	20	BA 20 (13)
BA 41	56, -36, 16	1.724	0.0009	12	BA 41 (11)
Lower FA associated with better performance (FA > test scores)					
(none)					

(a) Cluster group names assigned by SDM are illustrative and do not necessarily describe contiguous clusters.
(b) Voxel probability threshold: p = 0.005.
(c) Peak height threshold: z = 1.
(d) Cluster extent threshold: 10 voxels. Regions with fewer than 10 voxels are not included in the cluster breakdown.

Meta-Regressions and Sensitivity Analyses

The meta-regressions showed that the number of diffusion gradient directions had a significant moderating effect on the correlation between EDSS and FA in MS. Images acquired with more diffusion directions tended to have more significant clusters. None of the other meta-regressions for age, mean EDSS, magnet strength or lesion load were significant.

Table 7. Relevance of WM degradation in MS to EDSS score: meta regression analyses.

	MNI Coordinates	SDM Z-Value [a]	P-Value [b]	Number of Voxels [c]	Cluster Breakdown (Number of Voxels) [c]
EFFECTS OF MEAN AGE					
(none)					
EFFECTS OF MEAN EDSS SCORE					
(none)					
EFFECTS OF NUMBER OF DIFFUSION DIRECTIONS					
Greater FA-EDSS correlations specific to people with MS who were scanned with many diffusion directions (many diffusion directions > few diffusion directions)					
Posterior Cingulum	-18, -40, -4	2.380	0.000009	13	L BA 27 (11)

(a) Voxel probability threshold: p = 0.0005 for the slope and one intercept.
(b) Peak height threshold: z = 1.
(c) Cluster extent threshold: 10 voxels.

Table X. Significant associations between FA and MS diagnosis: sensitivity analysis. The table shows, by excluding each study from the meta-analysis one-by-one, which cluster groups were retained in the result. Each cluster group detected in the signed differential map is a column. A "Yes" indicates that the cluster in that column was still present in the result when re-running the meta-analysis with the study in that row excluded.

	Lower FA Associated with MS Diagnosis							Higher FA Associated with MS Diagnosis		
Excluded Study (author)	Corpus Callosum Body	Cerebellum	BA 10	BA 43	BA 9	BA 47	R Cerebellum Lobule VI	BA 22	R Posterior Internal Capsule	No. of Cluster Groups that Survived when Study was Excluded (*n* = 9)
Blaschek	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
Dineen	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	8
Liu	Yes	Yes	No	Yes	Yes	No	Yes	No	No	5
Llufriu	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
Onu	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
Schoonheim	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	8
Mazerolle	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
Hulst	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	8
Kern	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	9
Number of jack-knife analyses survived by the cluster group (*n* = 9)	9	9	6	9	9	8	9	7	8	

Table 8. Significant regional correlations between FA and EDSS scores in MS: sensitivity analysis. The table shows, by excluding each study from the meta-analysis one-by-one, which cluster groups were retained in the result. Each cluster group detected in the signed differential map is a column. A “Yes” indicates that the cluster in that column was still present in the result when re-running the meta-analysis with the study in that row excluded.

	Lower FA associated with greater disability	
Excluded study	L fornix crus / stria terminalis	Number of cluster groups surviving when excluding the study
Dineen	Yes	1 of 1
Giorgio	Yes	1 of 1
Liu	No	0 of 1
Smith	Yes	1 of 1
Yu	Yes	1 of 1
Kern	No	0 of 1
Number of jack-knife analyses survived by the cluster group	4 of 6	

Table 9. Significant regional correlations between FA and cognitive test scores in MS: sensitivity analysis. The table shows, by excluding each study from the meta-analysis one-by-one, which cluster groups were retained in the result. Each cluster group detected in the signed differential map is a column. A “Yes” indicates that the cluster in that column was still present in the result when re-running the meta-analysis with the study in that row excluded.

	Lower FA associated with poorer performance					
Excluded study	Fornix crus / stria terminalis	R posterior thalamic radiation	L posterior thalamic radiation	BA 20	BA 41	Number of cluster groups surviving when excluding the study
Dineen	Yes	No	Yes	Yes	Yes	4 of 5
Llufriu	Yes	Yes	Yes	No	Yes	4 of 5
Schoonheim	No	No	Yes	Yes	No	2 of 5
Yu	Yes	Yes	Yes	No	Yes	4 of 5
Hulst	No	No	Yes	No	Yes	2 of 5
Kern	Yes	Yes	Yes	Yes	No	4 of 5
Mazerolle	Yes	Yes	Yes	Yes	Yes	5 of 5
Number of jack-knife analyses survived by the cluster group	5 of 7	4 of 7	7 of 7	4 of 7	5 of 7	

5.4. Discussion

To my knowledge, this is the first voxelwise meta-analysis of studies relating tract FA to cognitive and physical disability in MS. The main findings were that the distributions for each were only minimally overlapping, and that reduced cerebral FA may have a greater relevance to cognitive than physical disability. This statistical neuroimaging consensus may aid in the future understanding of the underlying mechanisms for disability in MS.

While studies have previously shown that the posterior cingulum and splenium have lower FA in people with MS with greater physical disability [215, 218], the results suggest the association of further areas. Damage to NAWM resulting from the disease process and from Wallerian degeneration which may have gone undetected previously [236], reached significance in the analysis when the effects from multiple studies were combined. Another advantage of a meta-analysis is that the large combined sample size reduced the likelihood of type 1 errors. The results confirm the relationships between regions which are known to play roles in subcortical cognitive circuits like the thalami, fornices [35] and right posterior cingulum. The integrity of the posterior cingulum is affected in mild cognitive impairment and Alzheimer's disease [237], and is known to play a role in several relevant aspects of cognition including sustained attention and working memory [238]. The results confirm the importance of callosal tract injury to both cognitive and physical disability [220, 239, 240], with an anterior-posterior gradient suggesting a greater relevance of low FA in the genu and anterior body to cognitive

impairment and low FA in the posterior body and splenium being more relevant for physical impairments.

There are multiple factors to cognitive impairment and physical disability in MS mediated by widespread adaptive and injurious changes including grey matter and spinal cord damage. Given the assumption that reduced tract FA implies reduced connectivity between grey matter regions innervated by that tract, the meta-analysis indirectly supports the idea that disconnection of grey matter regions by damage to white matter plays an important role in contributing to symptoms, especially cognitive symptoms. Studies of disconnection of brain networks investigated by other means [163, 241] show results consistent with the structural disconnection phenomenon I have demonstrated. This suggests that cognition may be impacted by the functional connectivity of brain networks. For example, graph-theoretic studies of the MS brain have shown that reorganisation of brain networks is associated with disease diagnosis, disability and cognitive impairment [182, 242-245].

There are three limitations to this experiment. The main limitation is that, while the ES-SDM software uses a p-value threshold which the authors consider to be equivalent to the false discovery rate, the effect of false positives on this type of analysis is not well understood. Although spatial meta-analyses are known to suffer from false positive results [234], I took measures to account for this: I used the original t-statistic maps from each study instead of the reported coordinates and I reduced the number of comparisons performed by focussing only on white matter tracts. To further account for the possibility of false positives, I performed secondary analyses with a more conservative p-value threshold than is recommended, and showed that the main findings

retained significance. More work is needed in the future to assess the validity of the SDM method and the impact of false positives against other common methods for spatial meta-analysis such as activation likelihood estimation. Another major issue with SDM is that it offers no way to control the proportion of false positives at the cluster level, as opposed to the voxel-level. Compared to the main meta-analyses, the meta-regressions were inherently less powerful. They were based on mean values reported in each article rather than on the raw data; therefore, the results are likely less accurate and should be interpreted with care. Third, the meta-analysis of cognitively-relevant FA reduction included data from a variety of studies using a range of different cognitive tests. The meta-analysis is therefore based on the overlap of cognitive domains between these tests. The structure and spread of these data may have varied between types of test and therefore impacted the result.

Future work should explore the relevance of other DTI indices (MD, RD, AD) to functional impairment in MS. On one hand, FA is sensitive to ultrastuctural integrity and is established in the literature as a reliable measure of WM degradation [246]. On the other, there is evidence that FA does not sufficiently describe the underlying cellular changes [246]. FA is a summary measure, and is therefore less sensitive to the specific type of cellular damage. Several studies have given evidence that the AD, RD and MD scalars may be better predictors of physical disability scores in MS than FA, with studies mostly showing increased AD, MD and RD in specific regions in MS [211, 216-219, 225-227]. The correct interpretation of these measures is not clear, and are less frequently used compared to FA.

Finally, as the review was conducted in 2014, it is pertinent to now provide a brief update on any new studies which would be eligible for inclusion in the meta-analysis were it to be repeated. Six relevant articles have since been published [247-252], as identified using the same method described above. All six tested for differences in tract FA between MS and healthy groups, finding widespread reductions in FA throughout the TBSS skeleton specific to the MS group. No group findings conflicted with those reported previously or in the meta-analysis reported here. No articles provided voxelwise analyses of correlation of cognitive or physical disability scores to skeleton FA.

Conclusions

My results confirm that WM damage is widespread in MS and that the distributions of lowered FA in relation to cognitive impairment and physical disability are largely spatially distinct. The findings also suggest a greater relevance of cerebral WM damage to cognition than to physical disability, which may be more impacted by spinal cord damage.

The results of this study provide evidence for disconnection as a mechanism for symptomatology in MS. That disconnection is, in principle, a viable explanation for the occurrence of symptoms in MS, suggests that a graph-theoretic model of the MS brain may be a useful marker of cognitive impairment or fatigability. The evidence presented in this chapter serves to corroborate predicates 1 and 2.

6. Reliability of Brain Network Properties [2]

6.1. Introduction

With the recent rise in popularity of studies using graph theory, there has been a trend toward its application in clinical studies to investigate the mechanisms for emergence of symptoms from underlying neurologic injury or psychopathy. Such studies have demonstrated significant differences between MS and control groups in unitary global measures of network organisation (reviewed in section 4.3). Spurred by a wave of early positive findings, it has also been suggested by some authors that graph metrics may be adopted in future by clinical trials and in clinical practice as diagnostic tools [226, 254, 255]. It is therefore critically important that results based on graph analysis of MRI data are reliable.

Many studies have attempted to address the requirement for evidence of reliability by measuring the test-retest reliability of graph metrics in healthy people. In these studies, graphs were first constructed from data gathered at two or more timepoints and analysed for their organisational properties. Then, the extent of reliability between measurements over time was quantified using an intra-class correlation coefficient (ICC). Commonly, these studies also

[2] The systematic review component of this chapter has been previously published: 253. Welton, T., et al., *Reproducibility of Graph-Theoretic Brain Network Metrics: A Systematic Review.* Brain Connect, 2014. **5**(4): p. 193-202. My role in the publication was joint lead author with Mr Daniel Kent. We were jointly responsible for design, systematic review, data collection, analysis, manuscript preparation and submission. Other contributors to the publication were Dr Rob Dineen (concept and design) and Prof Dorothee Auer. All authors reviewed and commented on the final manuscript.

investigated how reproducibility was affected by various image processing and graph analysis strategies. While many of these studies find that reliability of graph metrics was good enough for future use in translational research, it is difficult to identify any consensus as to reliability in the general case and ideal approach to processing due to the wide range of methods and quality, as well as the presence of some conflicting results. One previous qualitative review of this literature tabulated parameters of the graph analysis but was not comprehensive and only included 6 studies [256]. Another review investigated reliability of resting-state fMRI measurements in human brain networks but did not focus on graph metrics [257].

A systematic review of literature on the reliability of graph metrics is therefore pertinent. Alongside these gaps in knowledge, the reliability of graph metrics between different scanners is unknown. This is an important prerequisite for applications in clinical trials and especially those which use MRI data from multiple sites or scanners. Additionally, while reliability of graph metrics in healthy people is well studied, there is a paucity of information on the reliability of graph metrics in disease. The experiments in this chapter aim to meet these needs. The results are discussed in the context of the underlying signal, which determines the connectivity of the graphs. If the MRI signal is not reliable, nor will the network metrics be, which are derived from it.

Reliability

Reliability is a broad term describing the extent to which a measure will produce the same result when performed multiple times. There is a range of terminology associated with the statistics of reliability, which is used variably throughout the literature, and often with overlapping or contrasting definitions;

therefore, a glossary is provided for reference in this chapter (table 10). Tests of reliability try to estimate the contribution of variability in measurement errors to variability in an observed measure, as opposed to the variability in the true score:

$$variance\ of\ observed\ score = variance\ of\ true\ score + variance\ of\ measurement\ error$$

Reliability is usually quantified by the ICC, of which there are two main types: consistency and agreement (figure 18). The two ICCs are defined as:

$$\mathrm{ICC(agreement)} = \frac{BMS - MSE}{BMS + (k-1)MSE + \frac{\boldsymbol{k(RMS - MSE)}}{\boldsymbol{n}}}$$

and

$$\mathrm{ICC(consistency)} = \frac{BMS - MSE}{BMS + (k-1)MSE}$$

where k is the number of observers, n is the number of subjects, BMS is the between-subjects mean squared, MSE is the mean squared error and RMS is the between-observers error mean squared. ICCs are generally interpreted as follows: <0.40, “poor”; 0.40-0.59, “fair”; 0.60-0.74, “good”; >0.74, “excellent” [258].

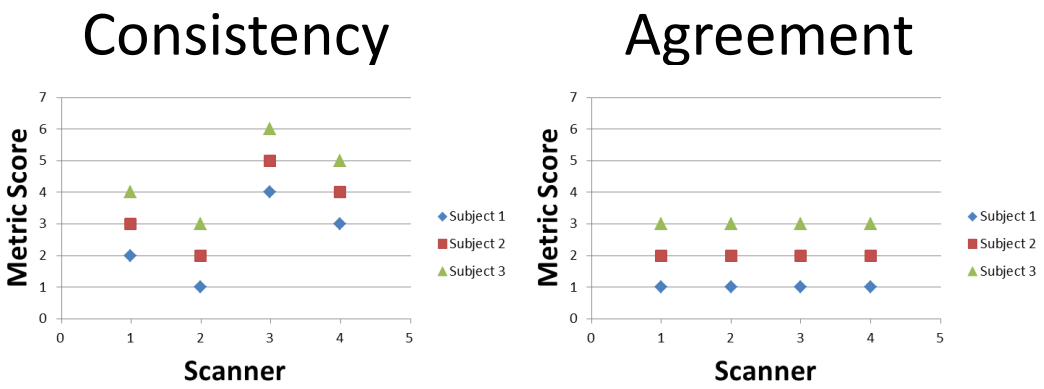


Figure 18. The two types of ICC. If a measure is consistent, it means that subjects' scores are consistent relative to one another, regardless of the absolute differences between observers (left). Agreement is a more stringent type of reliability in which the relative differences between observers matter (right). In this example, the observers are scanners and the measure is any graph metric. Measures which have good agreement are necessarily consistent.

Table 10. Glossary of common terms used in the statistics of reliability. Definitions from Bartlett and Frost [259] and McGraw and Wong [260].

Term	Definition
Reliability	An encompassing term describing the extent to which a test will produce the same result when performed multiple times. This may refer to either repeatability or reproducibility. Reliability is quantified by the intra-class correlation coefficients for either consistency or agreement.
Repeatability and reproducibility are subtypes of reliability that describe whether consecutive measurements are made in either identical *or* changing *conditions.*	
Repeatability	The repeatability of a measure refers to the amount of variation between successive measurements done on the same subject in *identical* conditions. Variability in repeated measurements is attributable only to the measurement process itself – the true value should be considered to be constant. Measurements should be made over a short period of time.
Reproducibility	The reproducibility of a measure refers to the amount of variation between successive measurements done on the same subject in *changing* conditions. Variability in reproduced measurements is attributable to the changing conditions, which could be the method, the observer, the setting or the period of time (which is long enough for non-negligible change to occur).
Consistency and agreement are both types of intra-class correlation coefficient that quantify reliability. Either statistic can be used to quantify either repeatability or reproducibility.	
Intra-class correlation coefficient (ICC)	The ICC is a statistic used to describe the resemblance of groups of units. An application of the statistic is in assessing consistency or agreement.
Consistency	In consistency measures, between-observer variance is considered irrelevant. Consistent measurements might vary in absolute value between observers but have low variance within observers. For example, the paired scores (1,3), (2,4), (3,5) are highly consistent but are not in agreement.
Agreement	In agreement measures, between-observer variance is considered relevant. Agreement is therefore more stringent than consistency because agreeing measurements also have a low variance between observers. For example, the paired scores (2,4), (2,4), (2,4) are both consistent and in agreement.

6.2. Aims and Hypotheses

This chapter is organised into three parts, corresponding to three experiments:

1. Reliability of graph metrics over time in healthy people. Through systematic review of the literature on reliability of graph metrics, I aimed to understand and summarise the current knowledge of reliability of graph metrics and the factors which influence reliability.
2. Reliability of graph metrics over time in people with MS. Using longitudinal data gathered from the BraNDy cohort (section 7.2), I aimed to test reliability of graph metrics in people with MS, and to compare it to that from previous studies in healthy people, as reviewed in the previous section. The null hypothesis for this experiment was that reliability of graph metrics in people with MS would be equal to that reported in systematically-identified studies of reliability in controls. I expected that reliability would be related to the severity of disease in the individual, the extent of subject head motion during image acquisition, the chosen graph density threshold and the resolution of the parcellation used.
3. Reliability of graph metrics between different scanners in healthy people. With a freely-available travelling-subjects dataset, I aimed to test how reliable graph metrics are between scanners, as opposed to over time. The null hypothesis for this experiment was that reliability between scanners would be the same as that measured over time in the same scanner.

The overarching aim in performing these experiments was to assess whether the reliability of graph metrics is sufficient for future applications in clinical studies. A secondary aim was to identify methodologic factors in graph analysis which are relevant to reliability and which therefore might be explored in order to improve it.

6.3 Systematic Review of Reproducibility of Brain Network Metrics

Methods

A literature search was performed independently by two researchers. The following four databases were searched through all languages and dates:

- MEDLINE (www.ncbi.nlm.nih.gov/pubmed)
- Web of Knowledge (wok.mimas.ac.uk)
- Google Scholar (scholar.google.co.uk) – only the top 100 results, sorted by relevance.
- OpenGrey (www.opengrey.eu)

The following search string was used:

(“graph theory” OR “graph theoretical”) AND
(“TRT” OR “test-retest” OR “reproducibility”)

Titles and abstracts of articles found with this search were first screened to identify those attempting to test the reliability of graph metrics in human brain networks. In a second screening, articles were excluded that did not meet all of the following criteria:

- The study must use only data from healthy human volunteers, as a disease process could confound the results.
- The study must measure reproducibility using either the ICC or coefficient of variation (CV) statistics, so that simple comparisons could be made.
- The article must include analysis of imaging data i.e. it must not be a review or meta-analysis.
- The full text must be available.

The reference lists of all remaining articles were searched for any other relevant results. From the remaining articles, I recorded the number of subjects, the type of scan, the inter-scan interval and the main conclusions about reliability. I also recorded from studies using diffusion tensor imaging tractography to construct graphs, the following: which software package was used for parcellation, registration diffusion modelling and fibre tracking, as well as the definition of edge weight used. All steps up to this point were performed independently by two researchers (myself and Dr Daniel Kent), and then a consensus was reached. Data from studies of functional and structural networks were tabulated separately, because I suspected that this would affect the graph metrics' reliability. I identified the following recurring themes in the literature and wrote for each one a qualitative synthesis of the included studies:

- Acquisition method
- Graph threshold type
- ROI size

- Preprocessing strategy
- Type of graph metric
- Fibre tracking algorithm
- Retest interval
- Type of ICC

Because most studies used custom software written in MATLAB with the Brain Connectivity Toolbox (BCT) and the formulae for the various graph metrics are well-defined, I assumed that the studies were equivalent in that respect.

I also applied a set of 10 criteria to assess the risk of bias of each study. The criteria were chosen based on previous quality checklists and were designed to place emphasis on methodological quality, rather than quality of reporting. Each criterion was given a weight of 1, 2 or 3 depending on its importance. Both researchers scored articles independently and then reached a consensus. The scores from the quality assessment were not used to exclude articles, but to determine how much influence they should carry in the review (table 11).

This review was originally conducted in August 2014 but updated in 2017 for this thesis using the same search strategy but filtering for articles published after August 2014.

Results

Literature Search

My initial search returned 393 results, 116 of which were duplicates. Of the remaining 277 articles, 244 were excluded at the first stage of screening because they did not aim to measure the reliability of graph metrics in brain networks in healthy people. Of the remaining 33 articles, one [261] was excluded because the authors analysed reliability of individual edge weights, not whole-brain summary metrics, two because the full text was not available [262, 263], three for not reporting ICC values [264-266], and one was excluded for analysing task data [267]. The 26 remaining articles were included in the review (figure 19). Three of the articles were added in the updated review.

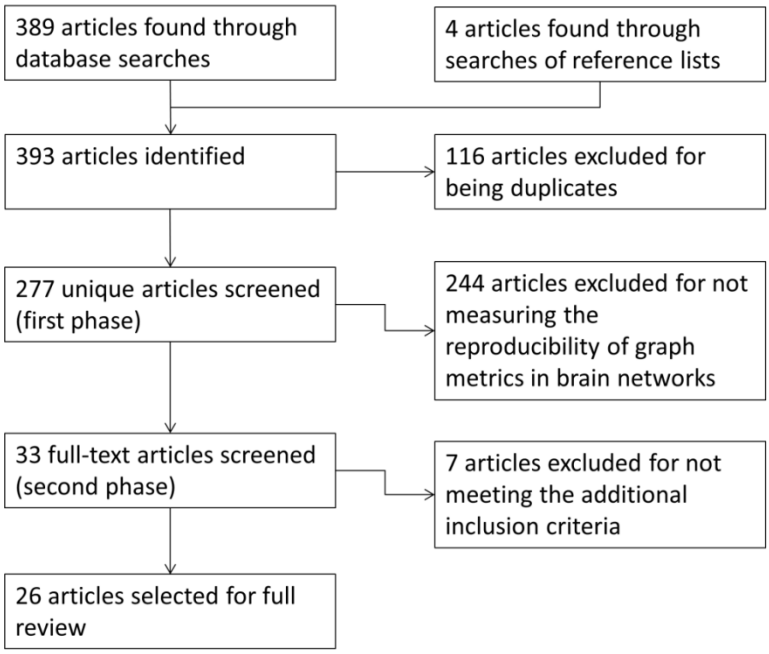


Figure 19. Flowchart showing the updated literature search process and numbers of remaining articles at each stage.

Risk of Bias

Table 11 shows the criteria and results of the quality assessment. The median quality score was 20 (the maximum possible score) and the range was 16-20. The reviewers (myself and Dr Daniel Kent) agreed on quality criteria scores in 196 of 230 total checks (89%; 10 criteria for each of 23 studies). The articles added in the update were checked for quality by myself only. Wherever there were conflicting scores, consensus was reached by discussion.

Table 11. The quality criteria and scores for each included article. ✓, the criterion was met; N, the criterion was not met or enough information was not recorded to make a fair judgement.

Study	Clear a priori hypothesis and design	Representative and appropriate sample	Appropriate scan sequences	Scanner strength >1.5T	All subjects received the same intervention using the same facilities	Conditions imposed on test subjects were unlikely to introduce bias	Appropriate retest interval	Appropriate data processing steps	Appropriate graph construction method	Appropriate statistics	**Total**
Andreotti	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Bassett	✓	N	✓	✓	✓	✓	✓	✓	✓	✓	20
Braun	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
Buchanan	✓	✓	N	✓	✓	✓	✓	✓	✓	✓	20
Cao	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
Cheng	✓	✓	✓	✓	✓	✓	N	✓	✓	✓	20
Dennis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	18
Deuker	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	20
Du	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Duda	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Fan	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Guo	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Jin	✓	N	✓	✓	✓	✓	✓	✓	✓	✓	18
Liang	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Liao	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Niu	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20

Owen	✓	✓	✓	✓	✓	✓	✓	✓	✓	N	18
Park	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Parker	✓	✓	N	✓	✓	✓	✓	✓	✓	N	16
Schwarz	✓	✓	✓	✓	✓	✓	N	✓	✓	✓	18
Telesford	✓	✓	N	✓	✓	N	✓	✓	✓	✓	16
Termenon	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Vaessen	✓	N	✓	✓	✓	✓	N	✓	✓	✓	16
Wang	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20
Weber	N	✓	✓	✓	✓	✓	✓	✓	✓	✓	17
Zhao	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	20

Study Characteristics

Table 12 shows a summary of the design and main findings of each included study. Study sample sizes ranged from 5-100 and numbered 676 in total. fMRI was the most frequent image acquisition method (13 instances; 50% of studies), followed by DTI (11 instances; 42%), MEG (2 instances; 7%), functional near infra-red spectroscopy (fNIRS; 1 instance; 4%) and arterial spin labelling (ASL; 1 instance; 4%). The interval between first and second scans ranged from being shorter than an hour to longer than a year. Some studies used more than one acquisition type (which is why the sum of acquisition methods above exceeds 26).

Of the 16 studies using functional data (including fMRI, MEG, fNIRS and ASL), at least one metric reached the “excellent” range in 11 studies [208, 268-275] [276] (64% of 15). At least one metric reached the “good” range in 4 studies [277-279] [280] (21% of 15). At least one metric reached the “fair” range in 1 study [281] (7% of 15). No studies reported only met the threshold for the “poor” range and one study did not fully report ICC data [282].

Of the 10 studies using structural data, at least one metric feel into the “excellent” range in 7 studies [256, 283-288] (70% of 10), the “good” range in 3 studies [289-291] (30% of 10), the “fair” and “poor” ranges in 0 studies.

Tables 13 and 14 list the highest ICC measurements reported in each study. Table 14 compares methods employed in studies of reliability in structural networks.

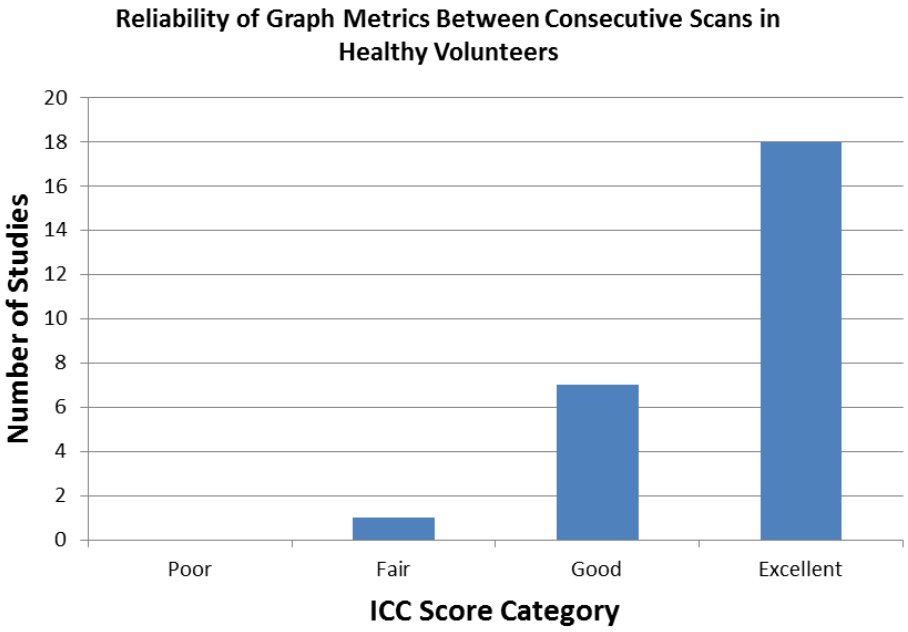


Figure 20. Number of studies in each category of ICC score, across all included studies. Studies were categorised based on the highest ICC reported in each, and therefore this chart represents the best-case scenario for each included study.

Table 12. Summary of included studies.

Study	N	Interval [a]	Acquisition [b]	Conclusions
Andreotti	19	<1 h	DTI	Smaller ROIs produced less reliable graph metrics.
Bassett	6	5.5 d	DTI, DSI	Metrics were more reliable based on DTI data compared to DSI data: DSI is too susceptible to noise. Metrics were more reliable based on structural atlases compared with functional atlases. Larger ROIs produced more reliable graph metrics. Correcting edge weights for ROI size did not improve reliability.
Braun	33	14 d	RS-fMRI	Reducing the length of the timeseries and global signal regression did not improve reliability. Using the first eigenvariate and the median timeseries instead of the mean did not improve reliability. Using a broader frequency range significantly improved reliability.
Buchanan	10	2-3 d	DTI	Networks based on white matter ROIs proved more reliable than grey matter ROIs. Using a deterministic or probabilistic tractography algorithm or varying the edge weight definition did not significantly affect reliability.
Cao	26	14 ± 2.1 d	RS-FMRI, n-back fMRI	Global metrics were more reliable than local metrics. Coarse structural atlases were less reliable than finer parcellation schemes.
Cheng	44	1 w	DTI	Reliability was sensitive to the two different edge weighting schemes tested.
Dennis	17	101 ± 18 d	DTI	Most metrics were reliable above densities of 0.3.
Deuker	16	4-6 w	RS-MEG, n-back MEG	Reliability was higher during performance of a task. Task practice was associated with reliability. Metrics were more reliable in low frequency bands.
Du	53	20 min, 6w	RS-fMRI	Reliability was sensitive to the scanning interval and order. Longer intervals had higher reliability. The metrics with the best reliability were the clustering coefficient, nodal degree and nodal efficiency.
Duda	21	<1 h	DTI	Choice of deterministic tracking algorithm did not significantly affect intra-subject reliability. Between subjects, some metrics are more sensitive than others to the choice of deterministic tracking algorithm. Choice of anatomical label set affects some metric scores and reliabilities.
Fan	16	1 y	RS-fMRI	Having eyes open or closed during rest affected reliability of graph metrics. Hubs were less sensitive to changes in rest condition.
Guo	24	13 ± 3 mon	RS-fMRI	Wavelet transformation of ROI timeseries improved reliability. Reliability of betweenness centrality was poor compared to degree and clustering coefficient. Soft- and proportional- thresholding did not improve reliability.
Jin	10	15 ± 8.4 d	RS-MEG	Reliability varies depending on eyes open and closed states, frequency band, metric type, and MEG sensor position. High frequencies were less reliable.
Liang	22, 25	<1 h, 11 ± 4 mon	RS-fMRI	Demonstration of differences in reliability with different image processing strategies (Pearson vs partial correlation, global signal regression and frequency band).
Liao	11	1 w	RS-fMRI	Degree centrality was more reliable at longer scan durations. Degree centrality was more reliable in some brain regions with a shorter TR. Global signal regression reduced reliabilities.
Niu	21	20 min	RS-fNIRS	Different haemoglobin concentration types produced different graph metric reliabilities. Denoising the data with ICA did not improve reliability.
Owen	10, 5	60.8 ± 33.6 d	DTI	Metrics were more reliable using individual segmentation than when using a high-resolution atlas. Number of

				streamlines per voxel and weighting edges by connection strength did not affect metrics' reliability.
Park	12	3 h (×8)	RS-fMRI	Local efficiency is more reliable than global efficiency.
Parker	28	<1 h	DTI	Demonstration of differences in reliability with different image processing strategies. Reliability of small-worldness index was low at low densities, compared to other metrics.
Schwarz	25	<1 h, 11 ± 4 mon	RS-fMRI	Reliability was higher over a short between-scan interval compared to a long interval. Deconvolution of white matter, CSF and motion timeseries increased reliability.
Telesford	45	<1 h	Executive task fMRI	Reliability was increased after spatial smoothing. Some metrics are more reliable in network hubs compared to non-hub nodes. Only the most stringent density thresholds significantly affected reproducibility. Mean degree was the least reliable global metric.
Termenon	100	"different days"	RS-fMRI	Metrics are reliable at scan durations over 7 min. Nodal metrics were most reliable in the DMN and the motor and visual networks.
Vaessen	6	14 ± 8 d	DTI	Number of diffusion directions and gradient amplitude had no effect on reliability.
Wang	25	<1 h, 11 ± 4 mon	RS-fMRI	Reliability was higher over a long between-scan interval compared to a short interval. Nodal metrics were more reliable than global metrics. Recommendations were given for graph construction depending on the type of atlas used.
Weber	22	30 min	ASL, risk/reward task fMRI	ASL perfusion fMRI produced more reliable graph metrics than BOLD fMRI. Reliability was lowest at low frequencies. Reliability was improved during task performance.
Zhao	24	>1 w	DTI	High resolution parcellation was more reliable than low resolution for global metrics, but low-resolution performed better for local metrics. Reproducibility of nodes in the primary cortex was better than in limbic/paralimbic regions. Hubs were more reliable than non-hubs. Multi-band DTI data were more reliable than conventional DTI.

[a] min, minutes; h, hours; d, days; w, weeks; mon, months; y, years.

[b] DTI, diffusion-tensor imaging; DSI, diffusion spectrum imaging; RS, resting state; fMRI, blood oxygenation level dependent functional magnetic resonance imaging; MEG, magnetoencephalography; fNIRS, functional near-infrared spectroscopy; ASL, arterial spin-labelling perfusion functional magnetic resonance imaging.

Table 13. Reported ICC values for various metrics in each study that reported exact ICCs from functional data. The studies not listed did not report exact ICC values. The highest values were chosen so that the values reported below represent the most reliable method tested within each study. These values could act as a guide for approximate expected values of ICC given different methodological choices.

Study	Sample Size	ICC Type	Acquisition	Clustering Coef	Char Path Length	Degree	Small-worldness	Glob Efficiency	Assortativity	Strength	Modularity	Loc Efficiency	Betweenness Cent	Hierarchy	Synchronisability	Diversity	Transitivity	Mutual Information	Cost Efficiency	Mean of all metrics
Cao	26	2,1	RS-fMRI, nback	0.27	0.54	0.38	0.57	0.67	0.71	0.59	0.58	0.25	0.28	-	-	0.43	0.69	-	-	-
Deuker	16	-	RS-MEG, nback	0.72	**0.90**	-	**0.77**	**0.87**	0.69	-	-	-	-	**0.83**	**0.76**	-	-	**0.87**	**0.81**	0.62
Du	53	2,1	RS-fMRI	0.69							0.45									
Braun	33	3,1	RS-fMRI	0.59	0.61	-	**0.76**	0.60	**0.75**	-	0.70	0.49	-	**0.75**	-	-	-	-	-	0.48
Niu	21	1,1	RS-fNIRS	**0.78**	-	**0.84**	-	**0.78**	-	-	-	**0.84**	0.68	-	-	-	-	-	-	0.71
Telesford	45	1,1	Task fMRI	**0.86**	**0.79**	0.29	-	**0.83**	-	-	-	**0.75**	-	-	-	-	-	-	-	**0.86**
Jin	10	3,1	RS-MEG	-	-	0.66	-	-	-	-	-	0.60	0.43	-	-	-	-	-	-	-
Guo	24	3,1	RS-fMRI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.63

Liao	11	1,1	RS-fMRI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0.78**
Median (Range)				0.71 (0.27 – 0.86)	0.70 (0.54 – 0.90)	0.52 (0.29 – 0.84)	**0.77 (0.57 – 0.76)**	**0.78 (0.60 – 0.87)**	0.71 (0.69 – 0.75)	0.59 (-)	0.58 (0.45 – 0.70)	0.60 (0.25 – 0.84)	0.43 (0.28 – 0.68)	**0.79 (0.75 – 0.83)**	**0.76 (-)**	0.43 (-)	0.69 (-)	**0.87 (-)**	**0.81 (-)**	0.67 (0.48 – 0.86)

Table 14. Reported ICC values for various metrics in each study that reported exact ICCs from structural data. The studies not listed did not report exact ICC values. The highest values were chosen so that the values reported below represent the most reliable method tested within each study. These values could act as a guide for approximate expected values of ICC given different methodological choices. As a rule-of-thumb, ICC scores are interpreted as follows: <0.40, poor; 0.40-0.59, fair; 0.60-0.74, good; >0.74, excellent (Fleiss et al. 2013) (shown in bold).

Study	Sample Size	Gradient Directions	ICC Type	Acquisition	Clustering Coef	Char Path Length	Degree	Small-worldness	Glob Efficiency	Assortativity	Strength	Modularity	Loc Efficiency	Nodal Distance	Betweenness Cent	Rent's Coef	Robustness Rand	Robustness Targ	Hierarchy	Synchronisability	Diversity	Mean of all metrics
Bassett	6	30	3,1	DTI, DSI	0.48	-	-	-	0.37	0.63	-	0.46	0.43	0.64	0.28	0.70	0.53	0.44	0.43	0.36	-	0.72
Andreotti	19	42	3,1	DTI	**0.93**	**0.91**	**0.90**	**0.91**	**0.92**	**0.87**	**0.87**	**0.91**	**0.93**	**0.76**	0.67	-	-	-	-	-	-	**0.86**
Cheng	44	48	3,1	DTI	0.54	0.28	-	0.55	0.64	-	-	0.67	-	-	-	-	-	-	-	-	0.70	-
Owen	10, 5	30	-	DTI	**0.88**	**0.94**	**0.90**	-	**0.94**	-	-	-	**0.87**	-	**0.94**	-	-	-	-	-	-	-
Dennis	17	94	1,1	DTI	0.58	0.59	-	0.52	0.60	-	-	0.61	-	-	-	-	-	-	-	-	-	-
Buchanan	10	64	3,1	DTI	**0.76**	0.64	0.66	-	-	-	**0.76**	-	-	-	-	-	-	-	-	-	-	-
Vaessen	6	32, 15, 6	3,1	DTI	**0.88**	**0.94**	**0.77**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Median (Range)					**0.76 (0.48 – 0.93)**	**0.78 (0.28 – 0.94)**	**0.84 (0.66 – 0.90)**	0.55 (0.52 – 0.91)	0.64 (0.37 – 0.94)	**0.75 (0.63 – 0.87)**	**0.82 (0.76 – 0.87)**	0.64 (0.46 – 0.91)	**0.87 (0.43 – 0.93)**	0.70 (0.64 – 0.76)	0.67 (0.28 – 0.94)	0.70 (-)	0.53 (-)	0.44 (-)	0.43 (-)	0.36 (-)	0.70 (-)	**0.79 (0.72 – 0.86)**

Table 15. Comparison of the approaches taken in studies of graph metrics' reliability in structural networks. AAL, Automated Anatomical Labelling; HOA, Harvard-Oxford Atlas; LPBA40, LONI Probabilistic Brain Atlas; WFUpick, Wake Forest University Pick; DKT31, Desikan-Killiany-Tourville; FSL, FMRIB Software Library; FACT, Fiber Assignment by Continuous Tracking; ANTs, Advanced Normalisation Tools; RK4, Fourth-order Runge-Kutta; TEND, Tensor Deflection; CATNAP, Coregistration Adjustment and Tensor Solving, A Nicely-Automated Program.

Study	Acquisition	Number of Diffusion Gradient Directions	Parcellation Scheme	Registration	Diffusion Modelling	Fibre Tracking Algorithm	Edge Weight Definition
Andreotti	DTI	42	FreeSurfer	FSL	FSL	FSL	[f]
Bassett	DTI, DSI	30	AAL, HOA, LPBA40, upsampling	FSL	TrackVis	TrackVis	[b, a]
Buchanan	DTI	64	FreeSurfer	FSL	FSL	FSL, FACT	[b, c, e]
Cheng	DTI	48	FreeSurfer	FSL	TrackVis	TrackVis	[c, g]
Dennis	DTI	94	FreeSurfer	FSL	FSL	FSL	[d]
Duda	DTI	34	DKT31	ANTs	CAMINO	FACT, Euler, RK4, TEND	[a]
Owen	DTI	30	FreeSurfer	FSL	FSL	FSL	[d]
Parker	DTI DTI	60	FreeSurfer, NiftySeg	FSL, NiftyReg	FSL, MRTrix	FSL, MRTrix	[a]
Vaessen	DTI	32, 15, 6	WFUpick	CATNAP	CAMINO	CAMINO	[a]
Zhao	DTI	128, multiband 4	AAL	CIVET	TrackVis	TrackVis	[a]

[a] The edge weight is given by the number of connecting streamlines between two ROIs.

[b] The edge weight is given by the sum of the connecting streamlines divided by the mean of the two ROIs' volumes.

[c] As (a), but correcting for streamline length.

[d] The number of fibres connecting the two ROIs normalised to the volume of the selected ROI.

[e] The mean FA value along interconnecting streamlines.

[f] As (c), divided by the sum number streamlines started from the ROIs, multiplied by the sum size of the two ROIs.

[g] 2 times the sum of the connecting streamlines between the two ROIs, divided by the sum volume of the 2 ROIs

Qualitative Synthesis of Results

Acquisition Method. Due to differences in the sensitivity to different properties of the brain between methods, the imaging acquisition type may be one of the largest factors influencing reliability. Of the metrics for which 3 or more ICCs were reported for both functional and structural studies, 6 of 7 were higher in metrics based on structural data). Considering the rapidly fluctuating haemodynamic state of the brain compared to its relatively static structure [292], this might be expected. None of the included studies drew comparison between reliabilities in more than one acquisition type.

Graph thresholds. Several different types of graph thresholding were used. Most commonly, a density threshold was applied. Other approaches were choosing a fixed threshold [282], mean degree thresholding [286], average path length [273] and calculating weighted variants of graph metrics [278]. No study has tried to test the ideal range of densities with which to threshold, but one study did compare fixed thresholding with “soft” and “proportional” threshold types, but found no significant differences in reliability [208].

ROI size. The effect of ROI size on reliability was investigated by three studies. The first tested the relationship between ROI size from within a single structural parcellation and reliability of nodal graph metrics over time [256]. A second study tested the reliability of metrics derived from three different anatomical atlases using DTI data to upsampled versions of the same atlases, by dividing each region into two [289]. Both of these studies found that, for DTI data, larger ROIs produced more reliable metrics than those based on more finely-grained parcellations. In contrast, a third study using fMRI data

found that metrics derived from a high-resolution functional atlas were more reliable than those from a lower-resolution atlas [277]. However, these results were based on ICCs from data averaged over three different tasks, rather than at rest. It was demonstrated that reproducibility in the three tasks varied significantly. It is not clear whether the reported difference in reliability between atlas resolutions would hold when compared at rest or within-task.

Preprocessing strategy. Different analysis strategies and tools were used to process imaging data in each study. One study did test two entirely different pipelines for analysis of DTI data and found significant differences in reliability between them, but was unable to show exactly which steps in the pipeline resulted in these differences [287]. In another study, the authors tested 7 different pipelines for fMRI processing, varying only one step in each and concluded that including a broader frequency band and using global signal regression produced the most reliable graph metrics [268]. A third study compared reliability after 5 different task regression methods, identifying two effective approaches [277].

Type of graph metric. Several classifications can be made of the many summary measures of graph organisation; for example, global and local metrics or weighted and binary metrics. Some of the included articles drew comparisons between reliability of metrics under different classifications. Two studies distinguished local from global metrics and found global metrics to be more reliable [256, 277]. Another found that first-order metrics were less reliable than second-order metrics (those derived from other metrics) [268]. Four studies investigated the relative reliability of individual metrics. One concluded that modularity was the most reliable [290], two more that the

clustering coefficient and global efficiency were the most reliable [271, 273], and another study found that degree was the least reliable [293].

Fibre tracking algorithm. Among the studies performing analysis of DTI data, eight different fibre tracking algorithms were used (table 14). In two of these studies, the reliability of metrics derived from graphs built using different fibre tracking algorithms were compared. In the first, the authors found that for any of the tested combinations of weighting or waypoint length threshold, neither the FDT (FSL Diffusion Toolbox) or FACT (Fiber Assignment with Continuous Tracking) algorithm was significantly better than the other [283]. The second study found no significant difference in reliability between four different fibre tracking algorithms [285].

Retest interval. Two studies investigated the effect of inter-scan interval on reliability of graph metrics and had divergent conclusions. In the first, it was found that reliability was better over a short interval [279]. The second study used the same publicly-available fMRI dataset, but found that graph metrics were more reliable over a long retest interval [274]. One difference between the two studies was that the first measured its long interval between scans >5 months apart, whereas the second measured it between the first scan (>5 months from the second) and the average of the second and third scans (<1 hour apart). The result could also have been confounded by the use of different parcellation schemes and removal of different sets of nuisance signals from the fMRI data.

ICC type. There are multiple types of ICC, each of which have a subtly different interpretation [294]. Amongst the reviewed studies, 9 used the

ICC(1,1), which is a measure of absolute agreement and is sensitive to differences in the mean score between observers [270, 271, 273-275, 279, 281, 282, 290]. One [277] used the ICC(2,1), in which observers are treated as random effects and emphasises the interchangeability of observers. Eight studies used the ICC(3,1), which treats observers as fixed effects and emphasises inter-rater consistency within a finite set of scanners (i.e. is not generalisable beyond that set of observers) [208, 256, 268, 278, 283, 285, 289, 295]. The coefficient of variation and Bland-Altman plots were also used to quantify reliability. None of the included studies compared ICC types or discussed the effect of their choice of statistical test on their results.

6.4. Analysis of Reproducibility of Graph Metrics in People with MS

Methods

This experiment used data from people with MS who had follow-up visits described, along with the sample characteristics, in section 7.2. The mean follow-up time was 46.6±14.8 days. After undergoing the preprocessing steps described for fMRI and DTI data described in section 7.2, and the steps for graph analysis described in section 7.2, I quantified reliability between the first and second visits using the ICC statistic (two-way random effects, single measures, consistency) and compared them to the ICCs reported in the systematic review of reliability of graph metrics in healthy people. I tested reliability under two parcellation schemes: Harvard-Oxford (126 regions) and FreeSurfer (164 regions). I also tested reliability separately at each density interval, as well as using the summary area-under-curve between the densities 0.2-0.5 as an overall measure. To identify factors contributing to

reproducibility, I tested for correlations between the between-visit difference in graph metrics and extent of motion, follow-up time, age and disease severity. I focussed on three graph metrics (the clustering coefficient, the characteristic path length and the global efficiency) because they are commonly reported throughout the literature, have good reliability and are easily interpreted. Focussing on a subset of important metrics also reduced the number of comparisons performed.

Results

Functional Connectivity

Eighteen fMRI datasets were included from 9 people with MS (72% female, mean age 41±9 years, mean follow-up time 46.6±14.8 days).

For the Harvard-Oxford atlas, reproducibility of graph metrics over time in people with MS over the range of densities 0.2-0.5 was graded as “good” for clustering (ICC=0.635) and global efficiency (ICC=0.659), and “fair” for characteristic path length (ICC=0.489), according to guidelines. For the FreeSurfer parcellation, clustering (ICC=0.499) and characteristic path length (ICC=0.573) were graded as “fair”, and global efficiency was graded as “good” (ICC=0.676).

The variation in reliability with different densities is shown in figure 22.

Table 16. ICC scores for consistency of graph metrics based on functional connectivity under two different parcellation schemes. An asterisk indicates ICCs in the “good” range. All others were in the “fair” range.

	Harvard-Oxford	FreeSurfer	Previous Studies (mean±SD)
Clustering Coefficient	0.635 *	0.499	0.644±0.231
Characteristic Path Length	0.489	0.573	0.710±0.164
Global Efficiency	0.659 *	0.676 *	0.750±0.112

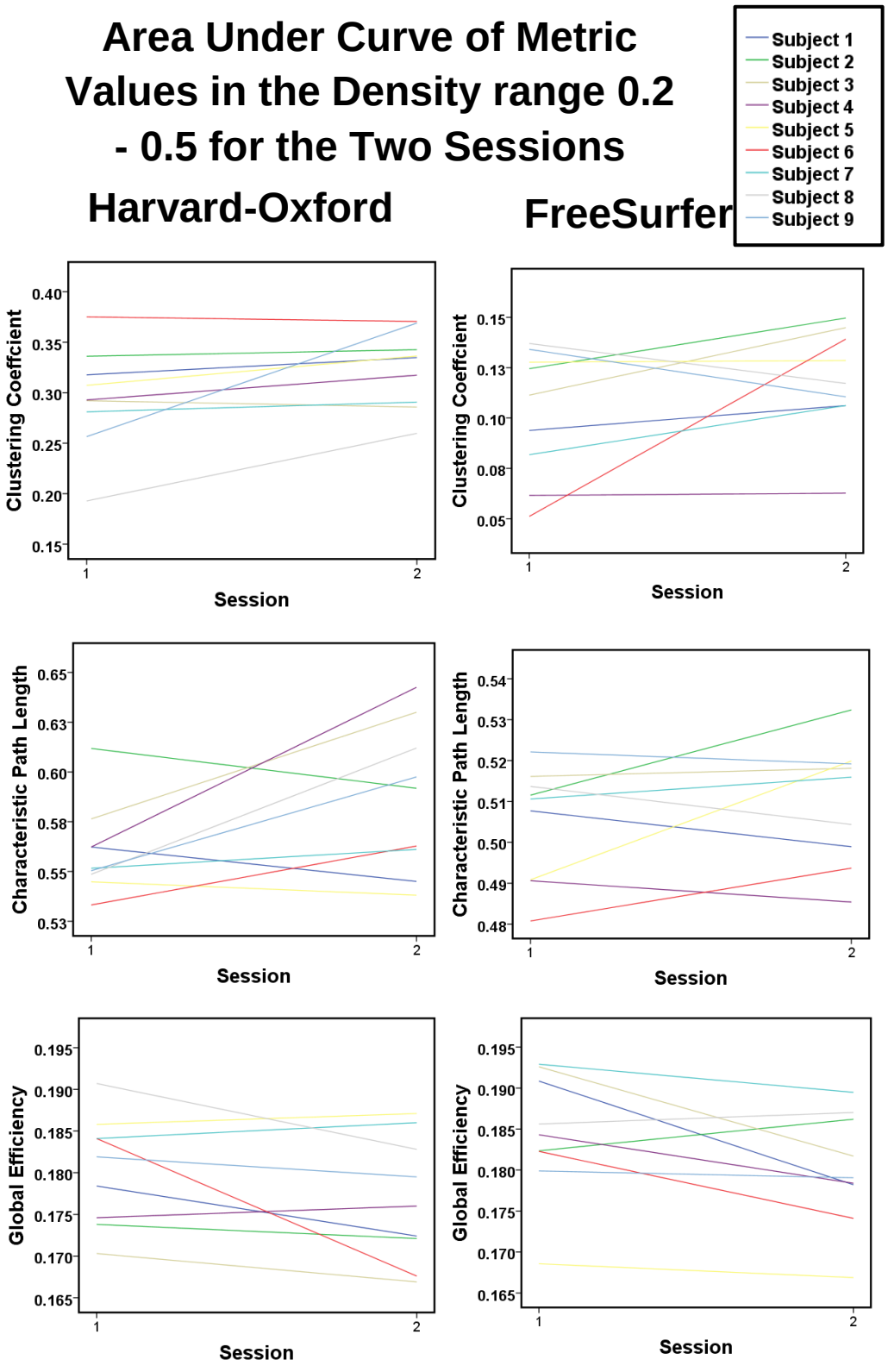


Figure 21. For the density range 0.2-0.5 only, the graphs show the change in metric score between the two sessions for the two parcellation schemes used.

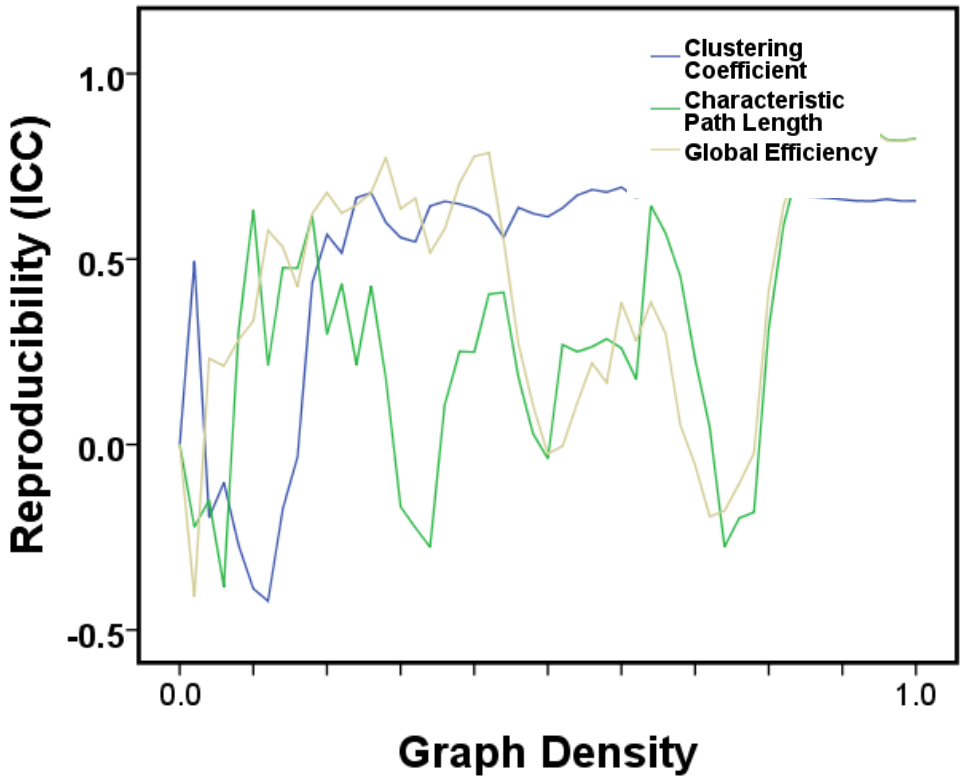


Figure 22. Reproducibility of each metric between the two sessions at different densities – Harvard-Oxford atlas only. The reproducibility of the metrics is low while the graph is very sparse and possibly fragmented but appears to become more stable at densities of 0.2 and above for the clustering metric.

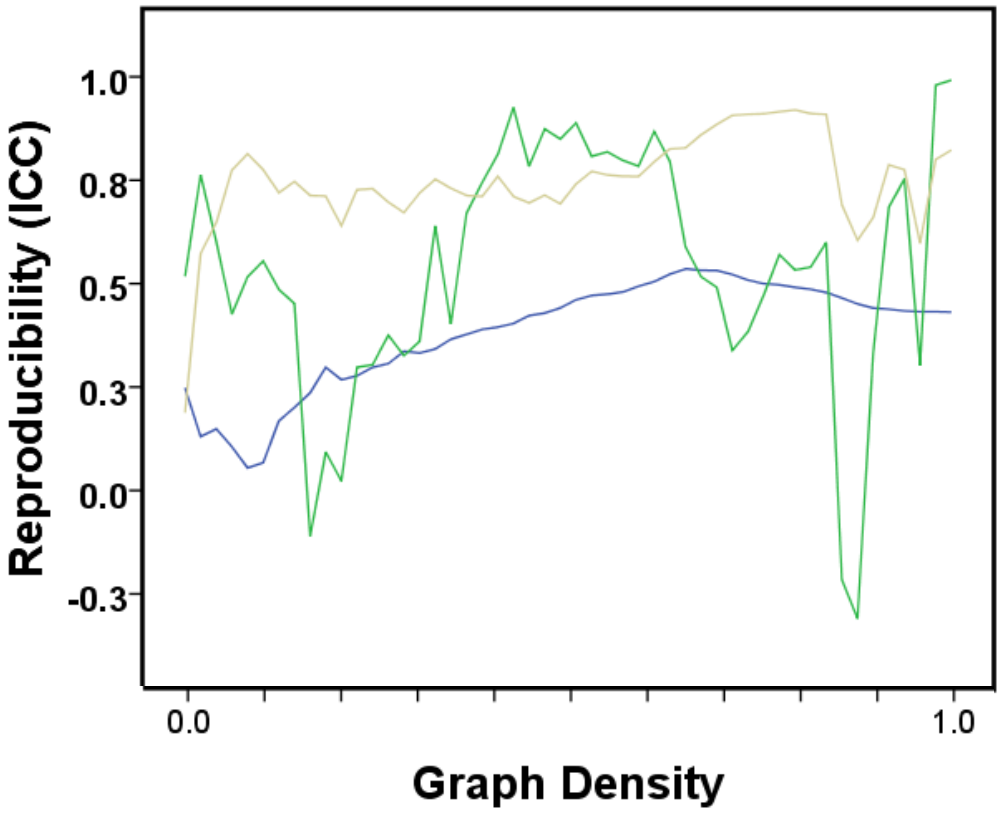


Figure 23. Reproducibility of each metric between the two sessions at different densities – FreeSurfer atlas only.

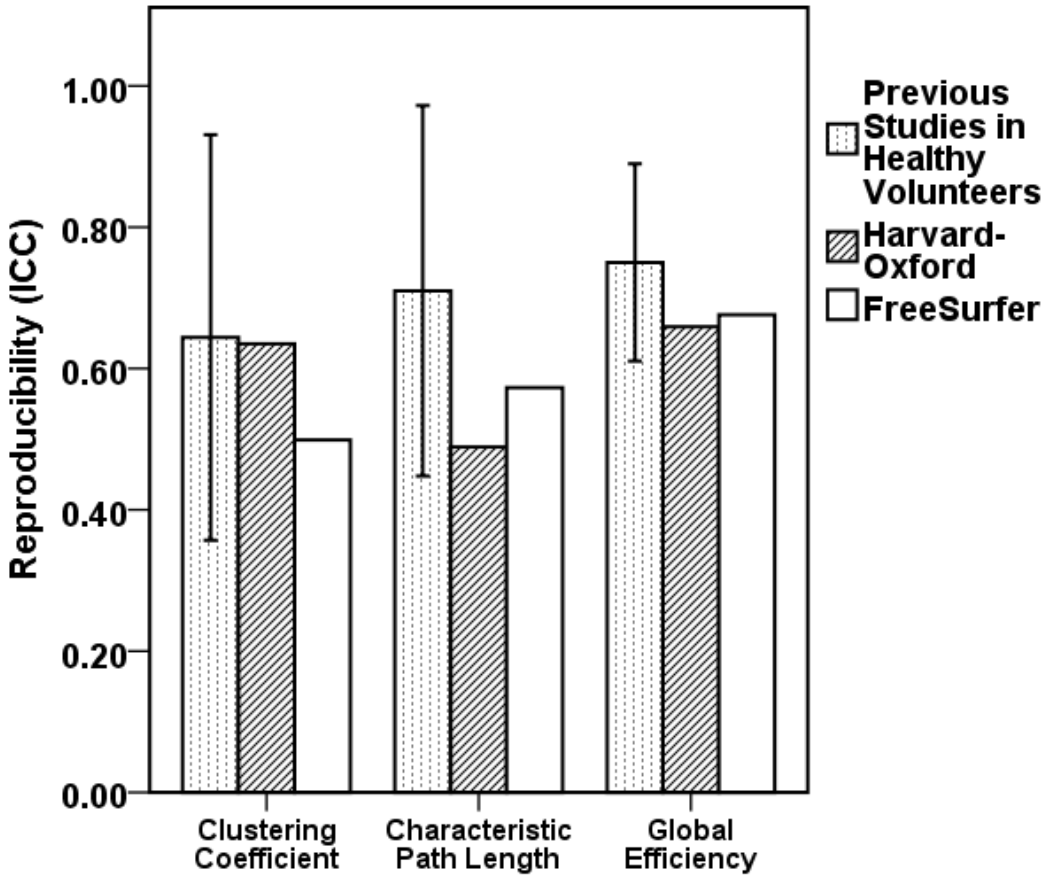


Figure 24. Bar chart comparing reproducibilities between graph metrics in the cohort of MS participants and in healthy volunteers from previously reported studies. A higher ICC reflects greater reproducibility. The bars for HOA and FreeSurfer data are from people with MS.

Compared to studies in healthy people reviewed in section 6.3 above, reproducibility of graph metrics in people with MS was only slightly lower in both atlases (median ICC across 5 studies; clustering ICC=0.723, efficiency ICC=0.788, path length ICC=0.704; figure 24).

For the HOA, none of the between-visit differences in graph metrics were explained by differences in disease severity, motion, follow-up time or age. MFIS (fatigue) score correlated with the between-visit difference in clustering coefficient (r=-0.792, p=0.11), although this did not survive correction for multiple comparisons and may be spurious given the sample size and

possibility of being driven by two outliers (figure 25A). For the FreeSurfer atlas, none of the between-visit differences in graph metrics were explained by differences in disease severity, motion or age, but the same correlation between difference in clustering and MFIS was significant (r=-0.863, p=0.0027; figure 25B) and, while it may also be spurious due to one outlier, passed Bonferroni correction (α=0.0033). The follow-up time also correlated with the difference in clustering coefficient (r=-0.706, p=0.033; figure 25C), but this was not significant after correction for multiple comparisons. However, these correlations may also be spurious due to the single outlier with a high MFIS score and large negative difference in clustering. When this outlier is removed, the above correlations are not significant. The cause for the outlier is not clear from demographic information but could be due to this individual sleeping on one occasion, although this was not verified.

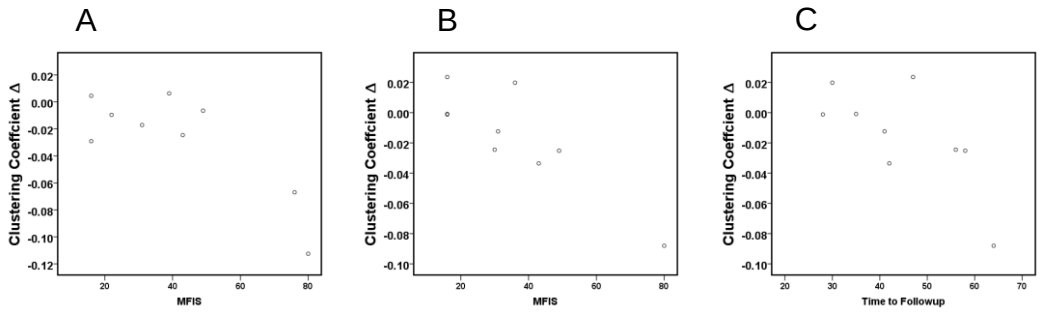


Figure 25. Significant correlations may have been driven by an outlier with a high MFIS score. (A) Correlation between MFIS and the inter-session change in clustering coefficient for the HOA. (B) Correlation between MFIS and the inter-session change in clustering coefficient for FreeSurfer. (C) Correlation between follow-up time and the inter-session change in clustering coefficient for FreeSurfer.

Streamline Density

Eighteen DTI datasets were included from the same 9 people with MS (70% female, mean age 41±9 years, mean follow-up time 46.6±14.8 days).

For the Harvard-Oxford atlas, reproducibility of graph metrics over time in people with MS over the density range 0.2-0.5 was graded as “good” for all three metrics (ICC=0.722, ICC=0.714, ICC=0.707). For the FreeSurfer parcellation, clustering (ICC=0.733) and global efficiency (ICC=0.689) were graded as “good”, and characteristic path length (ICC=0.800) was graded as “excellent”.

Table 17. ICC scores for consistency of graph metrics based on streamline density under two different parcellation schemes. An asterisk indicates ICCs in the “good” range. Two asterisks indicate ICCs in the “excellent” range.

	Harvard-Oxford	FreeSurfer	Previous Studies (mean±SD)
Clustering Coefficient	0.722 *	0.733 *	0.644±0.231
Characteristic Path Length	0.714 *	0.800 **	0.710±0.164
Global Efficiency	0.707 *	0.689 *	0.750±0.112

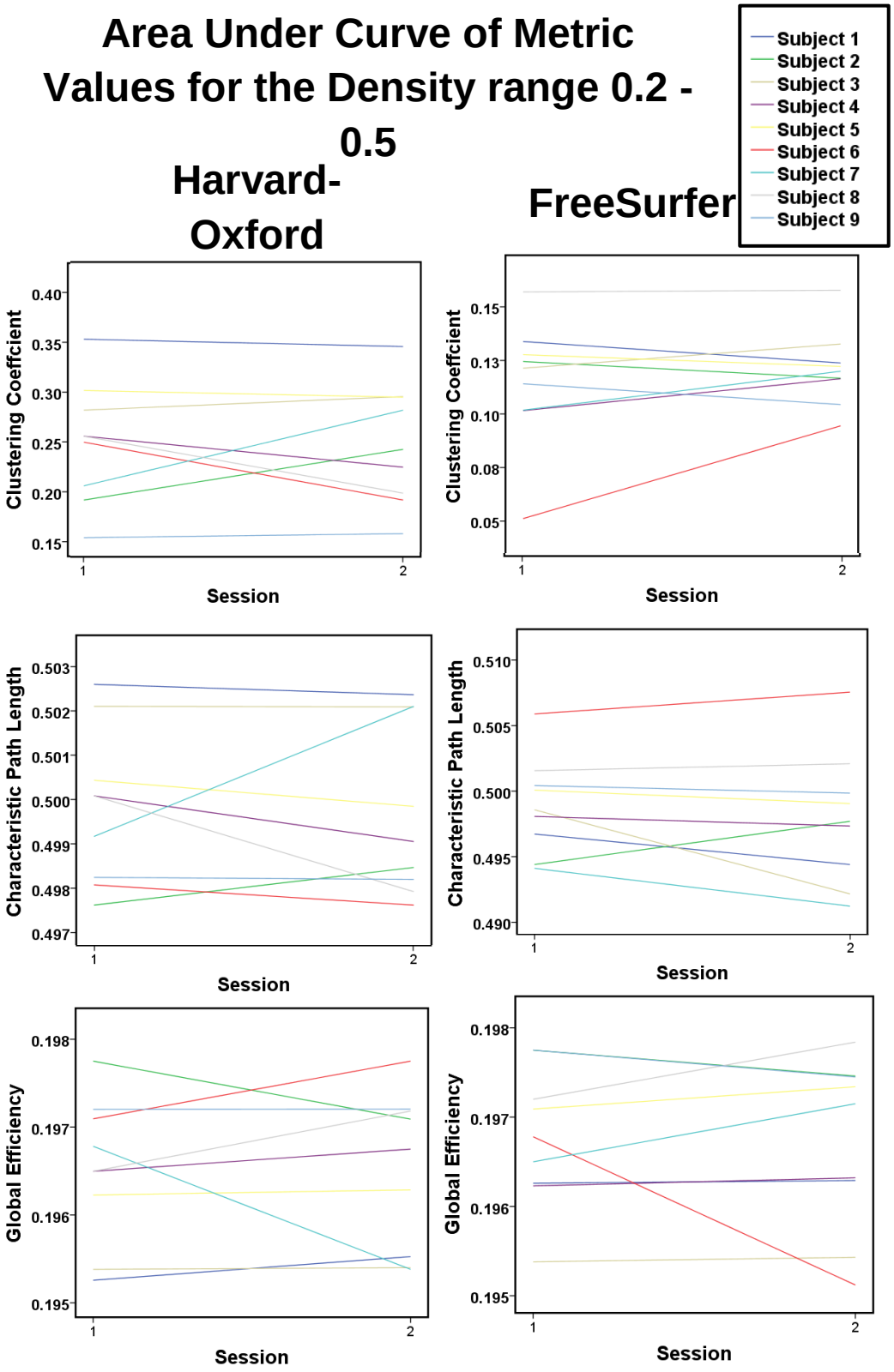


Figure 26. For the density range 0.2-0.5 only, the graphs show the change in metric score between the two sessions for the two parcellation schemes used based on DTI data.

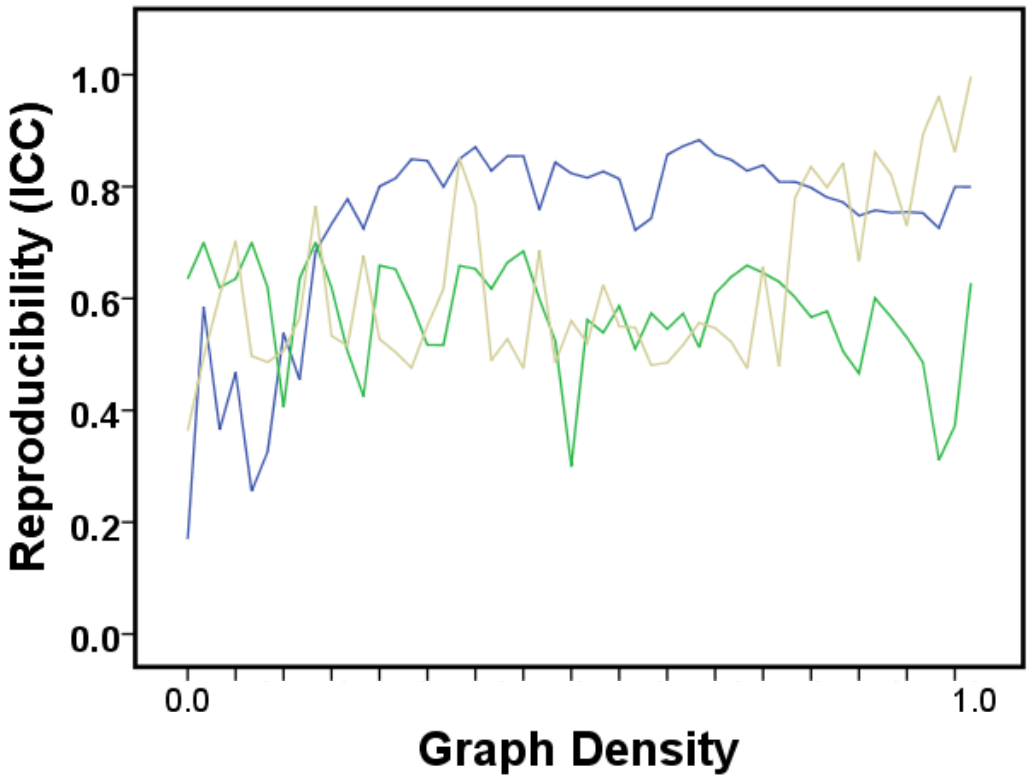


Figure 27. Reproducibility of each metric between the two sessions at different densities with DTI data – Harvard-Oxford atlas only.

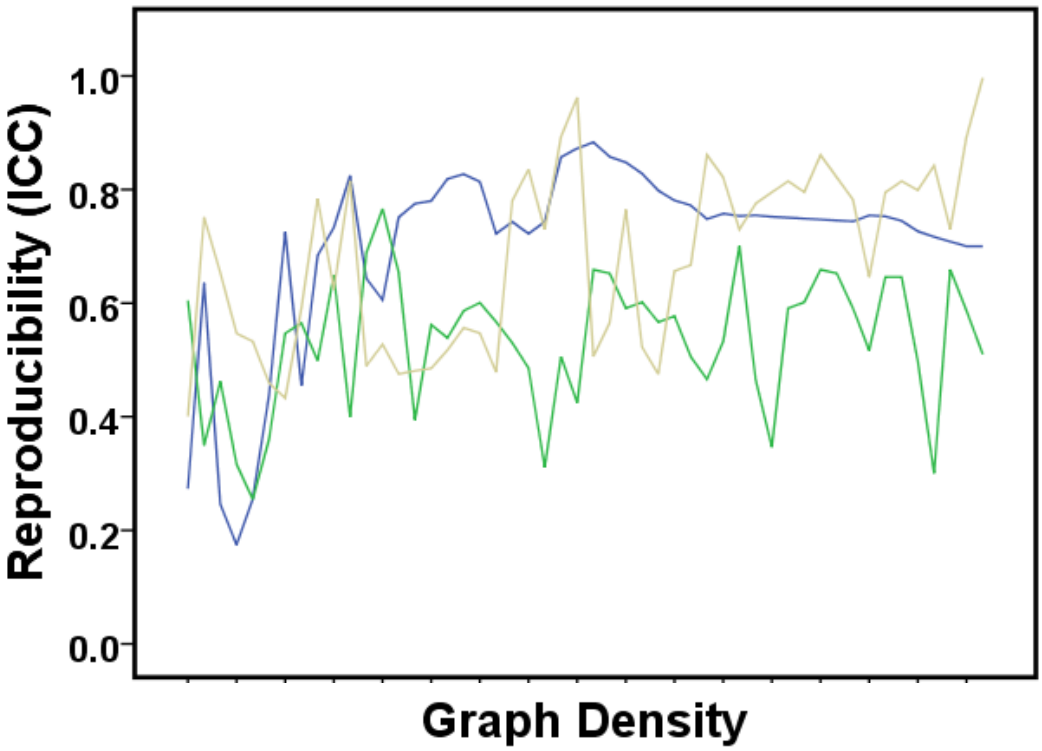


Figure 28. Reproducibility of each metric between the two sessions at different densities with DTI data – FreeSurfer atlas only.

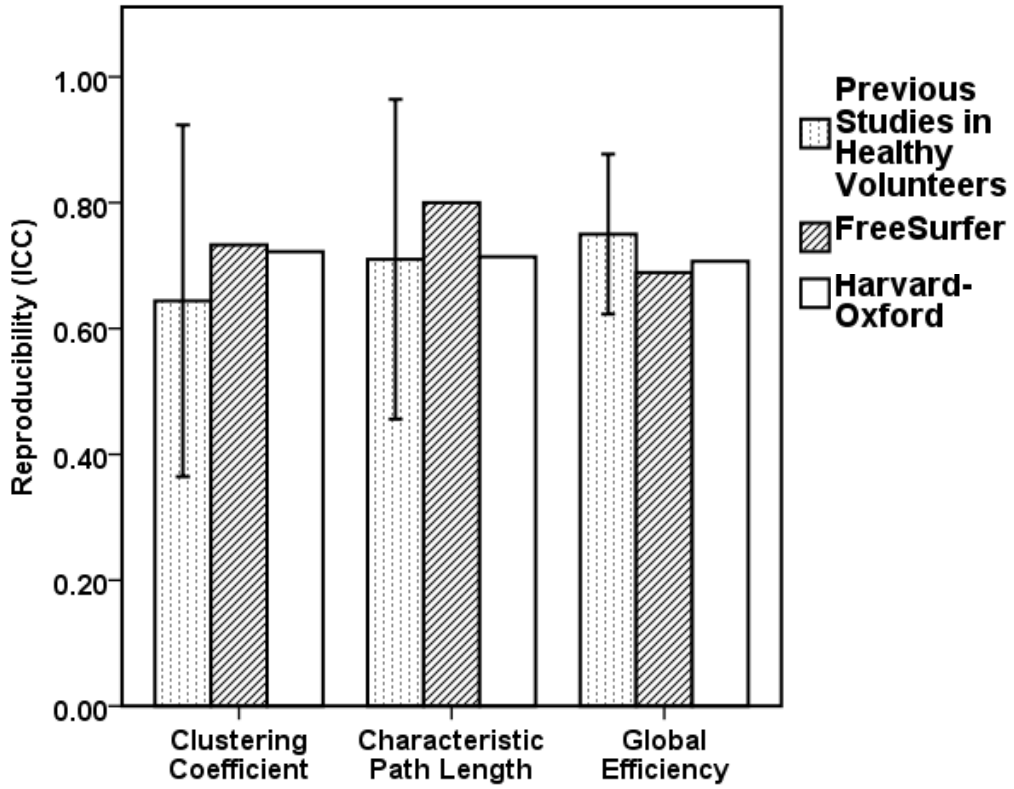


Figure 29. Bar chart comparing reproducibilities between graph metrics based on DTI data in the cohort of MS participants and in healthy volunteers from previously reported studies. A higher ICC reflects greater reproducibility. The bars for HOA and FreeSurfer data are from people with MS.

For the Harvard-Oxford atlas, follow-up time, age, motion, disease severity and MFIS did not correlate with any between-visit difference in graph metrics. For the FreeSurfer atlas, MFIS correlated with the between-visit difference in global efficiency (r=-0.717, p=0.030), but this was again likely driven by the outlier with a high MFIS score.

6.5. Analysis of Reliability of Graph Metrics between Scanners

Methods

The data used in this experiment were acquired as part of the Bioinformatics Research Network (BIRN) initiative and were downloaded from the Function

BIRN Data Repository (Project Accession Number 2007-BDR-6UHZ1). The particular dataset, “Phase I: Traveling Subjects Study”, includes data from 5 healthy subjects, who each travelled to 10 sites in the US. The test subjects were all right-handed males having at least 15 years’ education each. The median age was 25 and ages ranged from 20 to 28. Three were Caucasian and two were Kenyan African, and all spoke English natively. The subjects did not smoke, drink alcohol or caffeine and were free from psychiatric disorders.

During each site visit, test subjects received the same scanning protocol on two consecutive days. The scanning protocol included two resting-state fMRI scans, each lasting approximately 4.5 minutes (sequence: EPI or spiral GRE, plane: axial AC-PC, TR: 3000 ms, TE: 30 ms for 3T and 4T or 40ms for 1.5T, FA: 90°, voxel size: 3.475×3.475×4.000 mm contiguous, matrix size: 64×64×35 voxels in 1 shot, FOV: 22.24 cm, BW: ≥ ± 100 kHz, slice acquisition: interleaved). Other parameters (coil type, k-space trajectory, number of dummy volumes, manufacturer, model and magnet field strength) were not standardised; instead, investigators at each site were instructed to attempt their “best” acquisition with the means at their disposal. Table 1 details the attributes of the included scanners, and the parameters that varied between sites. Subjects were instructed to have their eyes closed.

Table 18. The included scanners and their attributes.

ID	BIRN Site Code	Coil Type	K-Space Trajectory	Magnet Field Strength (T)	Manufacturer, Model	Number of fMRI Volumes
1	3	TR quadrature head	Spiral	1.5	GE Nvi LX	85
2	3	TR quadrature head	Spiral	4.0	GE Nvi LX	85
3	5	GE TR research coil	EPI	3.0	GE	87
4	6	TR quadrature head	Dual-Echo EPI	3.0	Siemens Trio	85
5	8	TR quadrature head	EPI	1.5	Siemens Symphony	85
6	9	RO quadrature head	EPI	1.5	Philips/Picker	87
7	10	RO quadrature head	EPI	1.5	Siemens Sonata	85
8	11	Elliptical quadrature head	Spiral in/out	3.0	GE CV/NVi	87
9	12	TR quadrature head	EPI	1.5	GE Signa CV/i	87
10	13	TR quadrature head	EPI	3.0	Siemens	85

The following processing steps were performed using FSL version 5.0 [296]. Since the number of dummy volumes varied between sites, I removed the first 2 or 4 volumes such that each dataset had 83 volumes for further analysis. The generic processing steps were then followed for fMRI analysis, as in section 7.2.

Statistical testing was done using SPSS Statistics (version 21, IBM Corp, Armonk, NY, 2012). I quantified reliability with two main types of summary statistic. I used a two-way random-effects model to represent the data, meaning that the ‘observers’, i.e. scanners, were considered a random sample of a larger population and that each observer rates every subject. For the first test, I considered inter-observer variability to be relevant and so I used ICC(agreement) (known as ICC(2,1)). I also calculated ICC(consistency) to test how consistent measurements were with relation to one another across observers.

Results

Table 19 shows the inter-scanner reliability measurements for all five measured graph metrics, based on the first fMRI scan of the first visit for each scanner. Most ICCs were graded as “poor”, and only the consistency ICCs for the clustering coefficient and small-worldness were graded as “fair”. Figure 30 illustrates the metric scores from which the reliability measurements were made, and how they vary between scanners and subjects. Figure 31 illustrates variation between subjects and scanners.

Table 19. Values marked with an asterisk are in the “fair” range for ICCs.

	Clustering Coefficient	Characteristic Path Length	Small-Worldness	Global Efficiency	Modularity
Consistency	0.43 *	0.23	0.42 *	0.18	0.24
Agreement	0.31	0.25	0.36	0.19	0.25

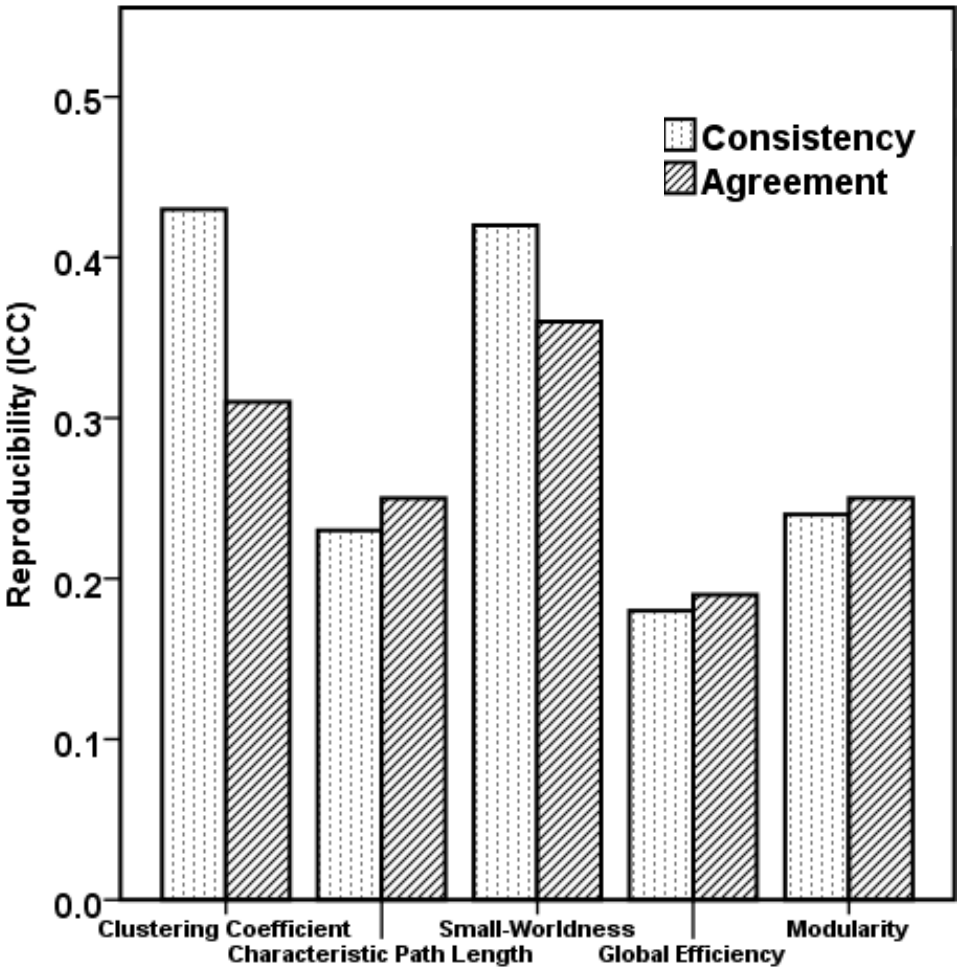


Figure 30. Comparison inter-scanner reproducibility between graph metrics.

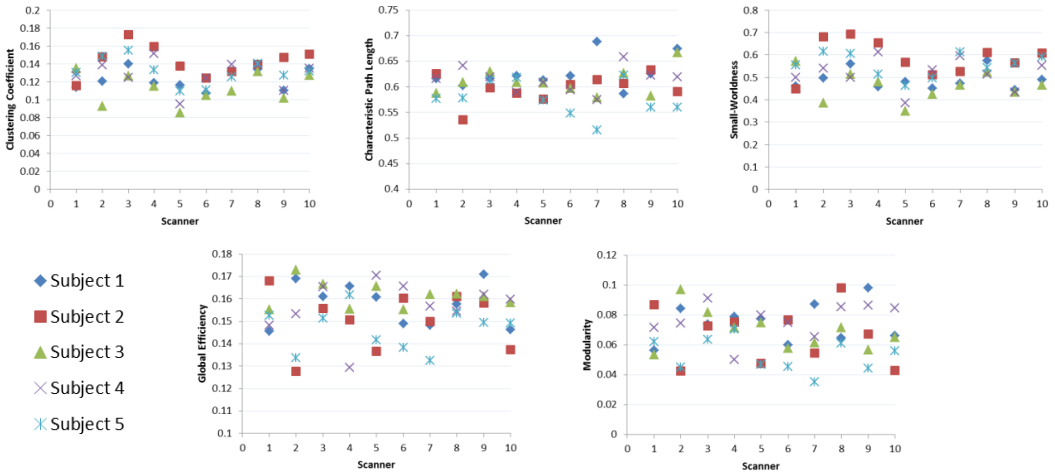


Figure 31. Dot plots of metric scores, based on individual fMRI scans. From left to right: (top row) clustering coefficient, characteristic path length, small-worldness, (bottom row) global efficiency and modularity. Dot colours indicate different subjects. That subject 2 tended to score highly and subject 3 tended to have a low score for both clustering and small-worldness suggests why those metrics had higher consistency scores than the others.

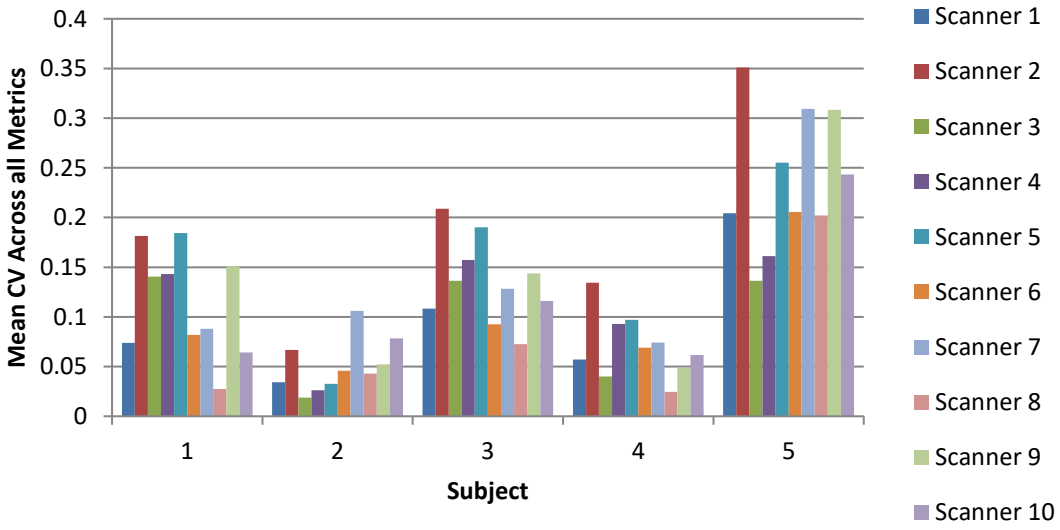


Figure 32. Line graph showing, for each subject, the mean CV (across all graph metrics) for each scanner. CV was measured across the four scans in a visit. Lower CV indicates better reliability. Visually, the graph suggests that some subjects tend to produce more reliable results than others, which may be due to variability in subjects' proclivity for motion.

Two-tailed paired-samples t-tests confirmed that CVs based on average metric scores were lower than those based on scores from a single scan, for both inter- and intra-scanner tests ($t=-5.018$, $p < 0.0001$ and $t=-5.480$, $p < 0.0001$, respectively). ICCs based on the average metric scores were not significantly higher, however (one-tailed sign test, $p = 0.4119$). Field strength was not related to reliability (1.5T vs 3.0T and 4.0T; $t=0.974$, $p=0.340$).

6.6. Discussion

Systematic Review

In this analysis I reviewed and summarised the published literature investigating reliability over time of graph theoretic brain network metrics in healthy people. I found that reported ICC scores were often in the ranges considered to be “good” or “excellent”, which suggests that reliability over time can be adequate in some cases. I found that scores varied between functional and structural networks; for example, for the metrics having 3 or more ICCs reported for both functional and structural networks, 6 of 7 were higher in the structural networks. Across all studies of functional networks, 6 metrics had median ICCs across 3 or more studies in the “good” or “excellent” ranges. In the studies of structural networks, 7 metrics met the same criteria. The most reproducible metrics were clustering coefficient, characteristic path length, global efficiency, degree, modularity, local efficiency and betweenness centrality.

I attempted to identify the methodologic factors most relevant for reliability. There was limited evidence that larger ROIs may be preferable for structural networks and smaller ROIs may be preferable for functional data. I found that

global metrics may be more reproducible than local metrics and second-order metrics may be more reproducible than first-order metrics. Reproducibility varied with acquisition type and depending on whether the subject was at rest or actively performing a task; for example, Wang, Zuo [274] show that, for resting-state fMRI data, degree was the most reproducible metric, whereas for Telesford, Morgan [273], under an executive task fMRI scan, degree was the least reproducible. There was also some limited evidence that the fibre-tracking algorithm used for DTI data processing had little effect on graph metrics' reproducibility, and that preprocessing steps taken can significantly alter metrics' reproducibility. The most optimal graph threshold type, retest interval and ICC type were not clear from the existing literature due to conflicting results, and the sample size and number of gradient directions had no clear correspondence to ICC scores.

My analysis of methodologic factors influencing reliability identified some important issues to be addressed by future research. First, the breadth of approaches and variability in the type of ICC used precluded meta-analysis and complicated the reaching of a consensus. For example, even where two studies used the same dataset and performed similar analyses, the results still varied drastically [274, 279]. There are still many unknowns in the applied methods, for example the ideal density threshold or range of thresholds, the necessary length of an fMRI scan, the type of atlas and the ROI size. Further, the most reproducible of these is not necessarily the most biologically plausible. Variability of study designs in the included studies meant that no image analysis pipeline stood-out as being the most reproducible. It is therefore recommended that future studies isolate one variable and study its

effect on reliability (as in [268, 277]), rather than varying multiple aspects of the pipeline. I also suggest that replication studies are performed to establish the generalisability of findings across cohorts and across more than one repeat scan. Four of the reliability datasets used are freely available to download [274, 279, 281, 283, 285].

The main limitation of the review was incomplete reporting in the reviewed articles. Meta-analysis was not possible without knowing the variances of the reported values. Meta-analysis would also be extremely limited due to the heterogeneity of methods employed which limited the only sensible method of synthesis to a qualitative discussion. Future studies should make all data available in future, including measures of spread so that future meta-analyses and meta-regressions can be undertaken.

Reproducibility of Graph Metrics in MS

The main aim of this experiment was to compare the reliability of graph metrics in people with MS to reliability in studies on healthy people reported in previous studies and reviewed in section 6.3. The main finding was that reliability in people with MS was not greatly different from healthy controls. This lends credence to their application in clinical studies and to the possibility that they may be useful in future as markers in clinical trials. It also suggests that disease status may not be a large factor in reproducibility, despite an assumed greater propensity for motion in diseased cohorts; for example, as a result of muscle spasms in people with MS. However, as a study of a relatively small cohort and the first of its kind, further similar studies should be performed in order to confirm or refute these findings.

A secondary aim was to compare reliability under two different parcellation schemes, the FreeSurfer and Harvard-Oxford atlases. On one hand, it may be expected that a network based on larger but fewer regions of interest and therefore based on averages over more voxels may be more reliable due to increased accuracy of the underlying connectivity measurements. On the other, it is arguable that a network comprising smaller but more nodes may be more resilient to changes in density, as there are more potential pathways between nodes and that the addition or removal of nodes might have less of an impact on network structure. It may also be important to consider that both atlases are based on sulcal depth [297, 298]. For this data, neither atlas produced more or less reliable graph metrics than the other, and the reliability of the metrics derived from both fell within the range of values reported in previous studies of reliability in healthy people, which were also based on a variety of types of atlases.

I also aimed to test how reproducibility varies with graph density. The traces for each metric over the range of densities follow similar patterns to those observed in previous studies [284]. My results imply that studies focussing on ranges below 0.25, such as that reported by Duda, Cook [285], may experience issues with poor reliability. My findings also highlight the importance of using a density range or range of raw thresholds due to the large possible variability in reliability with density in certain metrics. For the graphs of reliability over density, as edges are added incrementally, the point at which the graph becomes one single component is different between subjects, which may explain the large spikes and dips in reproducibility at low densities. The characteristic path length and global efficiency metrics are

much more variable and fluctuate in reproducibility greatly as the network changes in density, often following a similar pattern. While the initial dip in reproducibility is expected due to fragmentation of the graph, the dips around 0.3, 0.5 and 0.7 are less easily explained. They may be a result of large modules suddenly becoming connected as more edges are added. Alternatively, it may be that as all subjects tended to have very similar scores for characteristic path length and global efficiency at densities greater than 0.5, even small changes in the metric score can drastically change the rank order in comparison to the other subjects. This may also explain the peak at 0.9 density which is likely artefactual, where scores for some metrics tend to remain constant despite changing density. This is because at very high densities the whole graph is almost completely connected and likely has no relevance to the biological network. One explanation for this could be that the search resolution for the density threshold had a practical constraint, as increasing the resolution too high also raised the amount of time it took to run the analysis. Neurobiologically, high densities, such as those above 0.8, are unlikely to be relevant, as many of the included edges are of low (negligible) connection strength. Likewise, very low densities, are unlikely to be relevant as some connections with relatively high strength may still not be included in the graph. Interestingly, some peaks correspond between metrics, indicating points at which a small change shifted the network structure. The clustering coefficient is likely not as affected by this effect because it is an average of local clustering coefficients, rather than a true global metric. Second-order metrics are likely to be more susceptible to this effect. Using such analyses in future, studies will be able to quantitatively identify good ranges for reliability

and to use them for further analyses. However, this assumes that the most reliable range of densities also makes sense biologically, an assumption which remains to be tested.

I examined reproducibility of graphs based on two types of connectivity: functional connectivity and streamline density. First, the ICC scores for the streamline density data were higher than those for the functional connectivity for both the Harvard-Oxford and FreeSurfer atlases, and across all three metrics. This agrees with my findings in the systematic review of greater reproducibility in structural data. Having said that, the reliability scores for the functional data were also classed mostly as “good”, and all measurements in both datasets fell within the range of error from the results of the previous studies. A second issue of note is the similarity between the graphs for reproducibility over the range of densities. In both datasets the clustering metric tended to have the same shape: an initial dip followed by a steep rise and then remaining steady until the maximum density. The characteristic path length and global efficiency metrics did not appear to follow such a pattern and were much more variable, with large dips and peaks with changing densities. Interestingly, the two metrics’ dips and peaks often aligned, whereas the clustering did not at all. This is understandable, as the global efficiency metric is based partly on shortest path lengths.

Finally, I tested correlations of reproducibility with motion, follow-up time, age and disease severity. For both datasets, none of the between-visit differences in graph metrics were explained by these variables, suggesting that these differences were due to biological changes in network structure. For streamline density, only one correlation between the MFIS and between-visit

difference in global efficiency was significant. The same correlation with MFIS for functional connectivity was significant but all of these were driven by a single outlier with high MFIS score and hence are most likely not of interest. This could point to the poorly-controlled sleep state between sessions and between subjects as a confounding factor.

A limitation to this study was the potential for disease progression in the MS group. While the health state of the MS group may have changed only a little due to the relatively short inter-scan interval of 4-6 weeks, there is the possibility of change due to both treatment effects and disease progression. This may have caused the reproducibility of the data from the MS group to be lower than it would be otherwise. Second, there is a possibility of an increased propensity in the MS group to fall asleep, especially during the eyes-closed resting-state fMRI scan and in the more fatigued individuals. Again, this could have led to less reproducible data than in controls.

Reproducibility of Graph Metrics between Scanners

We found that reliability between scanners was generally poor, shown by the highest ICCs only reaching the “fair” range (clustering coefficient and small-worldness consistency ICCs were 0.43 and 0.42, respectively, and all other ICCs fell into the “poor” range: ICC < 0.40). Although this result remains to be replicated by other groups, it suggests that future multi-site clinical studies should investigate the inter-site reliability of graph metrics beforehand. This also suggests that some metrics, particularly clustering coefficient and small-worldness may be more reliable than others between scanners and therefore may be better candidates for use in future clinical neuroscience research. We also found that scanner field strength was not significantly associated with

reliability measures (there was no significant difference in either agreement or consistency ICCs for 1.5T scanners and 3.0/4.0T scanners). Our results also seem to show that individual subject differences may have an influence on the CV measurements, perhaps due to a propensity for some subjects to move more than others.

Ultimately, the reliability of functional brain network metrics depends on the reliability of the underlying BOLD signal. Friedman, Glover [299] tested, using a sensorimotor task fMRI paradigm from the same cohort, the inter-scanner reliability of percent signal change and contrast-to-noise ratio in 6 regions of interest. In agreement with our findings, they showed that reliability between scanners was low, with a median ICC for percent signal change between scanners of 0.22 and for contrast-to-noise ratio of 0.25. They go on to show how reliability can be improved by up to 123% when using average measures, dropping the sites with higher variability from the analysis, using larger regions of interest and regressing against image smoothness estimates between sites. The idea that fMRI measurements are more reliable over longer scans has previously been explored: Birn, Molloy [300] showed that inter-scan reliability could be greatly improved by increasing fMRI scan length from 5 to 13 minutes. Our relatively short resting-state scan times of 4.5 minutes may therefore partially explain the poor observed reliability. In another study, reliability across sites was measured as ICC=0.81 in the left dorsolateral prefrontal cortex and 0.95 for the left superior parietal cortex, indicating excellent reliability [301]. In test-retest reliability studies applying task vs rest paradigms (reviewed fully by Bennett and Miller [302]), the mean ICC was 0.50 and the range was 0.17 to 0.75. The propagation of error and

noise along the image processing pipeline, or use of poor quality data to begin with (known as the garbage-in-garbage-out phenomenon) is especially important when the final outcome is abstract or more highly derived, as is the case for graph metrics. It may therefore also be useful to investigate the reliability of the BOLD response to simple stimuli in people with MS relative to healthy controls. Studies testing the reliability of graph metrics in similar designs should also test the reliability of their fMRI data so that a relationship can be determined regarding whether or not global metrics are expected to be more or less reliable than the input data.

The disparity between the poor inter-scanner reliability and the good over-time reliability suggests that the poor reliability observed between scanners is inherent in the differences between scanners, rather than in biological variation reflecting some change over time in the subjects. The reason for this discrepancy is unlikely to be poor data quality, as that would also affect the within-scanner reliability. The suggestion is a systematic variation between scanners and therefore the issues with between-scanner reliability may be alleviated by applying the right standardisation procedures in preprocessing.

The main limitations to this experiment are, first, that certain parameters were not standardised across all sites. Operators were instructed to acquire the best images possible within the parameters given, but the field strength, coil type and k-space trajectory were not fixed. There are lines of evidence for each of these factors being important; for example, a 4.0 T magnet fMRI acquisition had 60-80% more significant voxels compared to a 1.5 T magnet, with a contrast-to-noise ratio 1.3 times higher [303]. A 32-channel head coil was shown to be better at detecting fMRI activation cortically compared to a

12-channel coil [304]. Second, the availability and sample sizes of this type of dataset are poor, which limits the potential for further study in this area. There is a need for better quality travelling-subjects datasets to address some of the concerns raised in this study.

Conclusions

I have shown that, with the correct processing steps, graph metrics have the potential to be reliable enough for future applications in clinical studies. However, the findings imply that graph analysis should not be employed “out-of-the-box”, because of the many factors influencing the reliability of results. Measured in the same sample across time, reproducibility can be excellent in both healthy and disease groups but, when measured across multiple sites, adequate reproducibility for application in clinical studies may be harder to achieve.

In comparison to other markers used in MS trials, the levels of reliability shown here are close to being adequate. A review of the psychometric properties of quality of life measures used in MS trials recommended that a minimum Cronbach’s Alpha coefficient of 0.70 is sufficient to establish reliability (Cronbach’s Alpha is conditionally equivalent to the consistency ICC) [305]. A study of the reliability of cognitive tests used in MS trials found that a composite measure including the PASAT, SDMT and others had an ICC of 0.90 and that the “learning and memory” and “ information processing speed” subcomponents had ICCs of 0.90 and 0.86, respectively.

7. Relationships between Cognitive Impairment and Brain Network Properties in Multiple Sclerosis

7.1. Introduction

In order to demonstrate statistical validity of graph-theoretic brain network metrics as markers of cognitive symptoms, it is necessary to show that they are related to disease status and cognitive performance. A number of previous studies have tried to address this aim, as were reviewed in section 4.3. Briefly: in people with MS who are cognitively impaired according to the PASAT, there is evidence for: (1) increased network segregation in the findings of increased clustering and modularity [184, 186, 187, 190, 192, 195] and (2) reduced integration in the findings of higher path lengths [184]. This implies a less efficient network structure which consists of more distinct groups of brain regions with reduced levels of communication between them [306]. This appears to be the consensus in most studies of cognition in MS using graph theory, and it intuitively makes sense that these changes would be associated with reduced performance of the relevant networks. However, there are some conflicting results suggesting greater network efficiency [185, 186] and, in two cases, no significant relationships [179, 183]. These negative and conflicting findings may be partly attributable to favourable adaptive plasticity, as two of the studied groups were CIS and one was only found in males, who are a minority among people with MS. Results suggest that these

whole-brain changes may be driven by changes in the frontal cortex and in the ventral stream/temporal cortex [189, 195, 196]. Such studies are fewer than 20 and further confirmation or refutation of their findings would be useful in understanding the processes underlying cognition in MS. Surprisingly, no study has focussed on the effects of brain network organisation on fatigue symptoms.

The broad aim of the work in this chapter was to provide evidence for the statistical validity of the brain graph metrics. This addresses the requirement for statistical validity described in research question 3. The further aims were to apply the graph-theoretic model of brain connectivity in the BraNDy cohort to derive summary graph metrics and investigate group differences and relationships to cognition and fatigue in MS.

Hypotheses and Aims

- There will be significant differences between MS and control groups in summary graph metrics describing brain network organisation.
- The strength and direction of these differences will be related to cognitive impairment and fatigue in people with MS.
- The observed differences will suggest increased network segregation and decreased network integration.

7.2. Generic Methods and Sample Characteristics of the BraNDy-MS Cohort

This section describes the methods used to gather and process the data used throughout the next sections. The BraNDy-MS study protocol described in this chapter was approved by a research ethics committee (Appendices B, C and

D). Data processing steps common to several experiments are given here and referred to in the respective sections. Processing steps which are specific to individual experiments are not described here.

Recruitment

Sample Size Estimates

I aimed to recruit 30 people with clinically-definite diagnoses of either RRMS or SPMS according to the McDonald criteria and 30 healthy controls. This target sample size was arrived at by a power analysis performed prior to the start of this study by Dr Rob Dineen based on an unpublished pilot study which included 8 people with MS and 8 controls using the G*Power software. Based on these data, in order to demonstrate a significant group difference in path length and small worldness at 80% power and with $\alpha=0.05$, for a one-tailed independent samples t-test, a sample of 27 in each group was required. The same sample size would be sufficient to show a significant group difference in clustering coefficient with 95% power and with $\alpha=0.01$. Acquisition of data from 3 additional participants per group was proposed to account for expected loss of data due to excessive motion or corruption (equivalent to an 11% expected rate of data loss). This gave us an estimated requirement of 30 participants per group.

I planned for 15 of the MS participants to return for a follow-up after 1 month. The sample size estimate of the follow-up group was partly based on a review by our group of previous studies which investigated the stability of network metrics in healthy people: in 23 studies with a similar design and all with positive findings, the mean sample size was 20.1 with a standard deviation of

10.8 [253]. I also based the estimate on another pilot study from our group (placebo-controlled cross-over) in which a graph-theoretic method was applied to analysis of network changes induced by administration of the sedative drug midazolam in 12 healthy volunteers. The study identified a significant increase in clustering coefficient after midazolam compared to placebo [307]. Given this information, I estimated that a sample of 15 subjects would be large enough to demonstrate variability in network organisation over time and to test whether it relates to changes in cognitive performance.

Eligibility

For the MS group, the eligibility criteria were as follows:

1. Aged 18-65 years. This age range was imposed for several reasons. First, to minimise the amount of discarded data due to motion which can be greater in children or the elderly. Second, to ensure that subjects had the capacity to give informed consent. Third, to avoid including subjects whose cognitive impairment may be related to other factors such as Alzheimer pathology. The maximum age corresponds to the UK age of retirement. People under the age of 18 were excluded because large variability in brain sizes could hinder quantitative brain image analysis, and because cognitive assessments could be skewed due to variability in level of education. MS in children often takes a more inflammatory form than MS in adults, which could complicate comparison with adult subjects [308]. A further argument is the natural prevalence and incidence of MS [2] which limits the age range of the population from which I could draw.

2. Diagnosis of clinically-definite RRMS or SPMS, defined by the McDonald criteria. RRMS and SPMS are often considered part of the same spectrum whereas PPMS and PRMS behave in a biologically different fashion. Additionally, PPMS and PRMS are much less common variants and therefore more difficult to recruit. These individuals are often in later stages of the disease and may be more susceptible to head motion.
3. Good comprehension of written English and fluent in English. This criterion was to ensure that the cognitive tests could be completed. Additionally, many of the cognitive assessments are only validated for fluent English speakers. Parts of the assessment are also based on response time.
4. Must be able to give informed consent. The subjects' ability to give informed consent was a judgement made by the researcher.
5. Must not be pregnant. This is to comply with MRI safety recommendations.
6. Must not have neurological, neurosurgical or psychiatric conditions other than MS. This was to ensure that any positive findings were due to MS, rather than any comorbid conditions.
7. Must not have relapsed within 30 days prior to the first visit or the period between first and second visits. This was checked so that any neuropsychological test scores or MRI outcomes were not degraded due to a temporary relapse or inflammation.
8. Must not have changed medication within 30 days prior to the first visit or the period between first and second visits.

9. For safety reasons, volunteers must not have any other contraindications to MRI, such as a pacemaker, replacement heart valve or claustrophobia.

The eligibility requirements for healthy subjects were the same, except for points 2, 6, 7 and 8.

To assess responsiveness, I had originally planned to scan 15 subjects with MS before and after a cognitive training intervention. Due to time constraints and difficulty recruiting, I instead co-recruited from a large MS trial which included a group memory rehabilitation (Cognitive Rehabilitation for Attention and Memory in people with Multiple Sclerosis; CRAMMS trial, ISRCTN09697576, CRN ID 17196, protocol paper: Lincoln, das Nair [309]).

Recruitment Protocol

I aimed to match the groups for age, sex and education (measured as the number of years' full-time education after completing secondary school). Volunteers were recruited via several routes:

- Posters placed on approved noticeboards around the school of medicine and in specific hospital areas where MS clinics and treatment take place. The poster design was ethically approved.
- Attendance at MS Clinics to explain any details of the study to potential volunteers referred by the doctor running the clinic.
- People with MS were referred from the CRAMMS trial if they agreed to take part.

- Other means including word of mouth, spouses and friends of participants or staff working at the hospital. Colleagues at the university were not recruited in order to avoid skewing the control sample toward a higher average education level in the control group.

MRI Acquisition

All subjects underwent the same protocol, regardless of which group they belonged to. MRI scanning was performed on a 3T Discovery 750 (GE, Chicago IL) at the Sir Peter Mansfield Imaging Centre in the Queen's Medical Centre Campus of the University of Nottingham. Before the MRI scan, all subjects completed a standard safety questionnaire. A pulse oximeter was attached to the subject's fingertip and respiratory bellows around their torso to record physiological parameters. The subject also held a sensor to communicate with the radiographer in one hand, and, in the other, a pair of buttons used to respond to the task. The MRI scanning session, including time to complete the required paperwork took approximately 1 hour altogether. The MRI protocol for second visits in the follow-up subgroup was the same. The MRI protocol comprised localiser and calibration scans followed by:

1. T1-weighted axial FSPGR (fast-spoiled gradient echo). FA=8°, matrix 256x256x156, voxel size 1x1x1mm, slice spacing 0mm, TE=3.17ms, TI=900ms, TR=8200ms, FOV=256mm, NEX=1, scan time 6:39.
2. ASL. Matrix 128x128x36, voxel size 1.88x1.88x4mm, TE=10.5ms, TI=1525ms, TR=4844ms, NEX=3, FOV=240mm, 3D acquisition, scan time 4:41.

3. T2 FLAIR. FA=111°, TE=120ms, TR=8000ms, TI=2250ms, matrix 512x512x46, voxel size 0.46x0.46x3mm, slice spacing 0mm, FOV=235.5mm, scan time 7:29.
4. DTI. Single-shot diffusion weighted EPI. Four b=0 volumes and 32 diffusion weighted volumes, matrix 128x128x66, voxel size 2x2x2mm, TE=minimum, scan time 4:56.
5. fMRI. FA=80°, TE=36ms, TR=2200ms, matrix 64x64x37, voxel size 3.75x3.75x3.6mm, 410 volumes, scan time 15:02.

The fMRI scan comprised two parts. The first part was in a resting state with eyes closed and lasted 7 minutes. After this time, the subject received a tap on the knee, after which they opened their eyes and viewed some instructions for the following task for 1 minute. After the minute, a task began and lasted 6 minutes. After the task, subject completed a final minute of rest with eyes closed. The purpose of the continuous scan combining rest and task was to allow us to observe functional changes during transitions from resting to active and from active to resting states. The task itself was designed to induce cognitive fatigue by sustained attention by Leavitt, Wylie [310]. It was modelled on the SDMT (described in section 7.2) but adapted for use in the MRI scanner to allow response using buttons (figure 33; mSDMT; as in [310]). I implemented the same task design in E-Prime (version 2.0; Psychology Software Tools, Inc.).

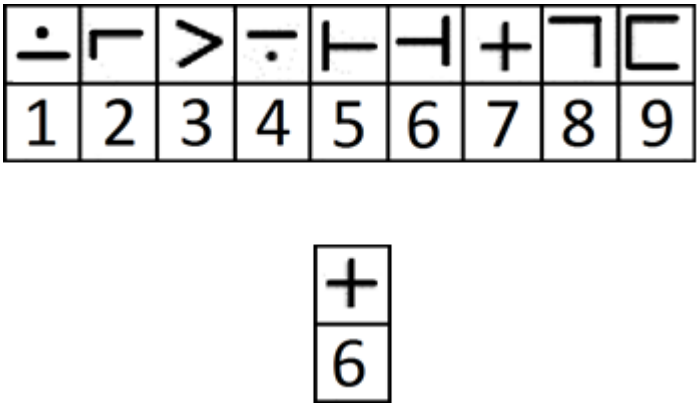


Figure 33. Example of the mSDMT stimulus. The mSDMT comprised 45 trials, each lasting 6 seconds with 2-second intervals (45*(6+2) = 360 seconds = 6 minutes). During each trial, the subject was presented with a row of 9 pairs of symbols and digits, and, below that, a single pair. The subject must decide whether the bottom pair matches any of the pairs in the top row and indicate their response by pressing one of two buttons under their finger (for match or no-match) on their dominant hand. After each trial, the symbols and digits were scrambled. Reaction time and whether the answer was correct or not was recorded for each trial.

Image data from the scanner were saved to DVD and then transferred to a Linux server for processing. The data were initially converted from DICOM to NIFTI format for storage using in-house software ("dtoa", Dr Paul Morgan, University of Nottingham, http://www.nottingham.ac.uk/~msapm1/software/). The CDs were used for long-term storage of the data within Radiological Sciences, but all further analysis was performed on the Linux server.

Neurologic and Cognitive Assessment

All subjects also had the same cognitive assessments, which occurred on the same day as the MRI scan. All tests were administered by myself. I was trained on how to administer the tests by a previous PhD student, Dr Afaf Elsarraj, and supervised by her for the first visit. The neuropsychological tests were chosen on the basis of providing measures which would capture a range

of possible cognitive deficits and fatigue, as well as the most likely confounds for performance, while fitting within a reasonable amount of time. The assessment took up to 1 hour altogether.

1. Multiple Sclerosis Functional Composite [311]. This test includes three components. The first is a nine-hole peg test (9HPT) in which the subject must add and remove pegs from holes in a board as fast as they can. The second motor component is a timed 25-foot walk (T25FW) to assess walking ability. The final component is a cognitive test of information processing speed, working memory and the PASAT 3-second version in which subjects hear a sequence of numbers and are required to say aloud the sum of the previous two numbers heard every 3 seconds [312]. To create the summary MSFC score, the Z-scores for the PASAT, the 9HPT and the 25FW are averaged. For the PASAT, the Z-score is based on the number of correct answers; for the 9HPT, the two times for each hand are averaged and then the reciprocals of the result for each hand are averaged, so that higher scores represent better performance; for the 25FW, the reciprocal of the average time of the two trials is used. The test Z-scores are first calculated without reference to normative data to allow for comparison between the two sessions. Comparison to normative data is also made to assess the generalisability of the results using the mean and standard deviation of subscores from over 5000 people with MS pooled across 14 trials [311].
2. Beck Depression Inventory 2nd edition (BDI-II) [313]. As a known confound of clinical outcomes in MS [314, 315], I measured depression

so that I could later test and potentially regress it against the main outcomes. This questionnaire is widely used and comprised 21 questions. The score is out of a maximum 63.

3. Modified Fatigue Impact Scale [316]. The MFIS questionnaire measures the impact of fatigue in terms of physical, cognitive and psychosocial functioning. The score is the sum of answers to the 21 questions (maximum 63).
4. Pittsburgh Sleep Quality Index (PSQI) [317]. Sleep is often disturbed in MS due to range of possible causes may be related to fatigue symptoms [318, 319]. The PSQI is a quick but reliable general measure of sleep quality. There are 19 items which produce 7 subcomponents or one global score. I used the global score only.
5. Symbol Digit Modalities Test [104]. The SDMT is a substitution task in which subjects are required to fill in numbers in boxes which match certain symbols. The subject must fill as many boxes as possible correctly within 90 seconds. The score is the number of boxes filled correctly.
6. Neurological Fatigue Index [24]. This 23-item questionnaire assesses four aspects of neurological fatigue: physical, cognitive, relief by diurnal sleep and abnormal nocturnal sleep. The maximum score is 69.
7. Attention Network Test (ANT) [320]. The attention network test measures reaction times to quantify the efficiency of alerting, orienting and executive function networks. Subjects must respond with left or right buttons corresponding to the direction of a centre arrow in a row of five arrows.

Lesion Load Quantification

The total volume of lesions present on the T2 FLAIR images was measured for each subject using the JIM software package (v5.0, Xinapse Systems, Essex, UK). The software tool features a semi-automated lesion segmentation algorithm. I was supervised in this task by Dr Rob Dineen and Dr Afaf Elsarraj. First the user identifies a set of coordinates which are central in MS lesions (figure 34). These coordinates are used as input to the algorithm, which uses a detailed prior knowledge of lesion distribution and size to form a probability template based on a large sample of people with MS. The tool takes the seed coordinates and dilates the region of interest around them until a threshold is reached, taking into account the lesion probability map. This occurs in 3D between slices, so seed coordinates only needed to be placed once per ROI, rather than per ROI and per slice. In some cases where the generated ROI was too large or too small, the “weight on prior probabilities” option was tuned. ROIs can be manually edited if necessary. Using this software allowed my segmentations to be more reliable and less dependent on the operator. When the ROIs have been created, the software calculates the total number of ROIs and their volume in ml.

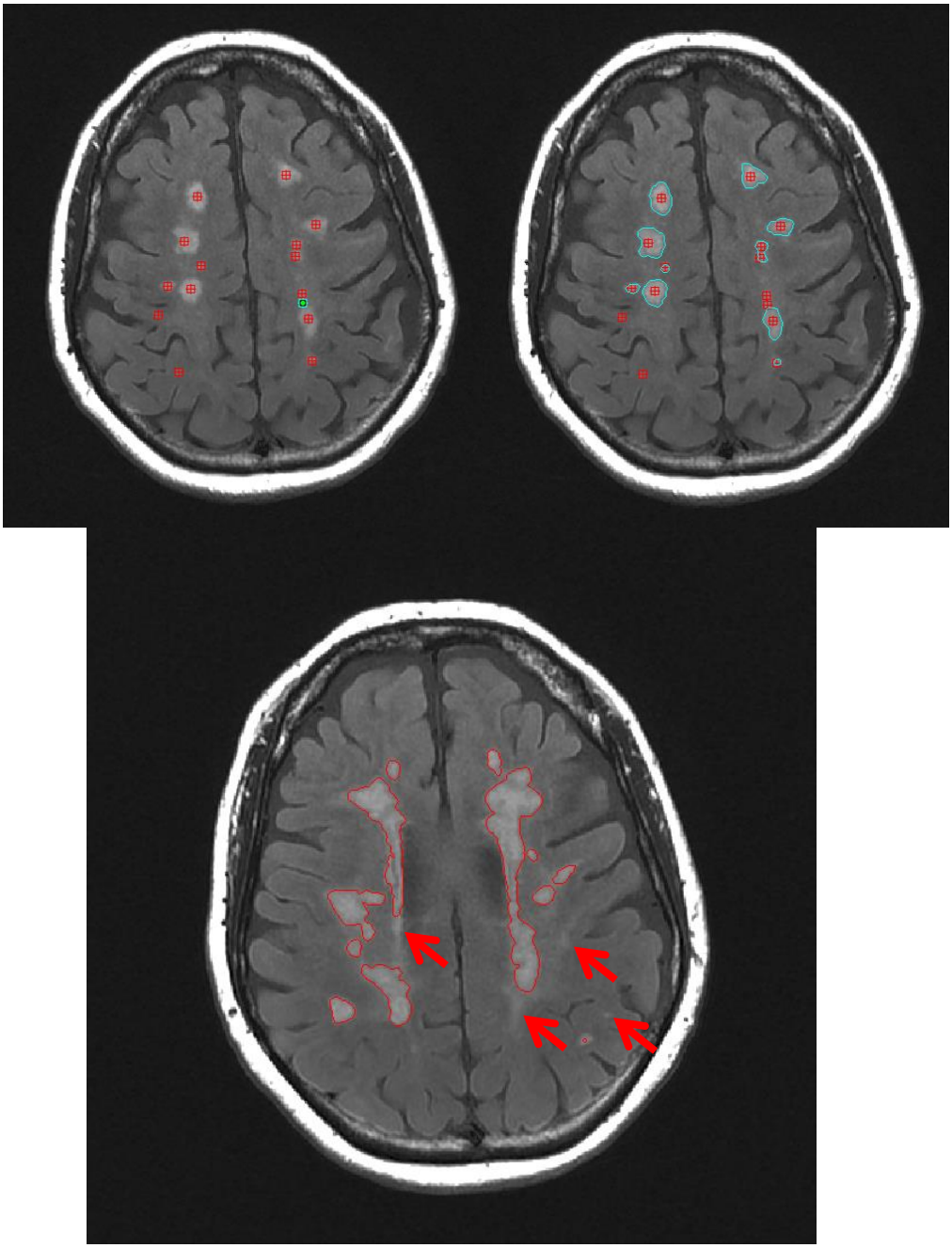


Figure 34. An example of the output from the semi-automated lesion segmentation on subject 36. The top left image shows the coordinates identified as the centre of lesions. The top right shows the same slice after automatic lesion segmentation. The lower image shows some ROIs generated by the software in a case where the threshold was set too low or more seed coordinates needed to be added. It is clear that the software does not detect diffuse damage or very small lesions with the chosen threshold in the lower image (red arrows). In this instance, the threshold would be altered to allow larger ROIs and more seed coordinates would be placed.

Brain Segmentation

Brain segmentation serves two purposes in the experiments: First, it allows us to measure atrophy, parenchymal brain volume (the sum of grey and white matter volumes) and parenchymal brain fraction (the ratio of parenchymal brain volume to total brain volume including CSF). The parenchymal fraction is useful as a measure of atrophy because it can be measured without longitudinal data. Second, it is used for the definition of sets of regions of interest which can be used in analyses of connectivity. In all segmentations, raw volumes are always normalised to the volume of the whole brain to account for intersubject differences in brain size and therefore to allow fair comparison. Segmentations were performed on the T1 data because of its high resolution and good contrast between grey and white matter and CSF. Three types of segmentation were performed for each dataset.

The first aimed to measure whole-brain, white matter, grey matter and CSF volumes. The tools used for this were FSL Brain Extraction Tool (BET; v2.1) [321] and FSL FAST (v4.0) [322]. BET applies a deformable model which is the adapted to fit to the brain's surface to estimate its volume. Once the inputs and outputs are specified, the only option which may need changing is the "fractional intensity threshold". For every dataset, BET was run with an initial threshold of 0.5 and the resulting volume checked against the input T1 image to verify that it worked well. If the estimated volume was too low or too high, the threshold was adjusted until the outline of the estimated extracted brain volume looked correct. The best threshold varied between subjects but was generally adjusted to 0.4 for the best results. The algorithm for the FAST segmentation of grey matter, white matter and CSF, is based on a hidden

Markov random-field model and is fully automated and didn't require any corrections, although there are options to input tissue means or prior probability maps to improve the accuracy. The accuracy of all outputs was verified by overlaying them on the T1 image. When the 3D volumes are generated, the fslstats command line was used to calculate the volume of the ROIs based on the voxel size.

The second type of segmentation aimed to generate a set of 164 cortical and subcortical grey matter ROIs for connectivity analysis using the FreeSurfer (v5.3.0; http://surfer.nmr.harvard.edu/) [323] software package and the Destrieux atlas [298] from the MRI data for each individual. FreeSurfer is a large software package which includes tools for cortical segmentation and uses a model of the cortical surface based on probabilistic information from a manually labelled training dataset. This geometric information and neuroanatomical convention. The Destrieux atlas is based on a scheme which divides the cortex into gyri and sulci by measuring curvature of the surface. The command used to run the segmentation, "recon-all" requires the input data to be organised into folders and numbered, but takes as its input the T1-weighted image, which must be converted into MGZ format. The command took approximately 11 hours to run per subject on the Linux server. The main output of the program is a 3D volume of cortical ROIs, each labelled with a different intensity. A separate file contains the subcortical ROIs in the same format. These outputs are checked against the input image to verify their accuracy.

In the third type of segmentation, I applied the Harvard-Oxford atlas (HOA) [324]. The HOA includes 63 cortical and subcortical ROIs based on gyri. In

contrast to the FreeSurfer method, in this method I applied the pre-existing atlas directly, rather than generating tailored regions of interest for every individual. Because the HOA regions are bilateral, they were first split in two, for left and right hemispheres using the fslroi command. This yielded an atlas with 126 ROIs.

fMRI Pre-processing

fMRI preprocessing was done using tools in the FMRIB Software Library (FSL; v5.0.7; http://www.ndcn.ox.ac.uk/divisions/fmrib/) [296]. The first two volumes (4.4 seconds) of each dataset were removed to allow for the brain tissue to reach a steady state of radiofrequency excitation. The start of the resting-state, transition, task and final rest of the scan were at 0, 420, 480 and 840 seconds and volumes 0, 191, 218 and 382, respectively. 4D images were cut-off at these points where required for individual experiments using the fslroi command. Head motion was corrected for by affine registration to volume 96 (the centre of the resting-state) using MCFLIRT. Datasets were checked for excessive motion (defined as a cumulative rotation of >3° or translation >4mm, approximately corresponding to the voxel size and being a common but arbitrary standard among MRI studies). BET was applied to remove non-brain structures. Spatial smoothing was not applied in order to avoid introducing local correlations between adjacent ROIs. The FreeSurfer brain images were registered to the fMRI datasets using FSL FLIRT and the resulting transformation applied to the ROI volumes in order for them to be mapped to fMRI space. The fslmeants command was used to extract the mean BOLD signal within each region of interest over the timecourse. Using MATLAB (version R2013a, The MathWorks, Natick, MA), the linear trend was

removed from the extracted timeseries to account for scanner drift. The following variables were then regressed from the timeseries: three translation and three rotation parameters, the mean signal from the ventricles, and the mean signal from the WM. The WM and ventricle ROIs were from the FAST segmentation described above, but eroded using a 4x4x4 voxel kernel to ensure that no partial volumes were included. These regressors were decided upon because they represent a good balance between denoising and signal loss for datasets of a similar number of timepoints as ours [325]. Finally, to create association matrices, functional connectivity between every pair of ROIs was quantified by the Pearson correlation coefficient.

DTI Pre-processing

DTI data were processed using FMRIB's Diffusion Toolbox [326]. First, eddy-current induced distortions and subject motion were corrected using the eddy program. All image volumes were registered to the average b=0 image by affine transformation. BET was applied to remove non-brain structures. BESTPOSTX was then used to estimate diffusion parameters. ROIs from the two atlases were moved to diffusion space by registering the FreeSurfer brain image to the average b=0 image by affine transformation and applying the transformation matrix to the ROIs. The streamline tractography algorithm, PROBTRACKX2, was used for fibre tracking, with 5000 streamlines seeded per voxel. For each voxel in a given seed mask, the algorithm propagates streamlines along the path with the largest principal axis of diffusion for a fixed length or until termination criteria are met. The termination criteria in the analysis was set as the streamline reaching the voxels in a termination mask. This process was applied to every possible pair of ROIS in each atlas to

determine matrices of the number of reconstructed streamlines between them. This was measured twice (once for each direction) because the number of streamlines between seed and terminal ROIs could vary depending on direction. This meant that for streamline data, adjacency matrices were asymmetric. The following formula was then applied, which sums the streamline counts for each direction and scales the result by their volumes, as in [283, 284, 286]:

$$a_{ij} = \frac{s_{ij} + s_{ji}}{m_i + m_j}$$

Where a_{ij} is the corrected summary streamline count, s_{ij} and s_{ji} are streamline counts in each direction between ROIs i and j, and m_i and m_j are voxel counts in each ROI.

Sample Characteristics

A total of 60 volunteers were recruited between 8th September 2014 and 26th January 2017. This comprised 37 people with MS and 23 control subjects. A summary of the basic demographic data for the cohort is shown in tables 20 (MS) and 21 (controls). The 30-30 split was deviated from because I also co-recruited from the CRAMMS trial. Figure 35 shows the recruitment pathways.

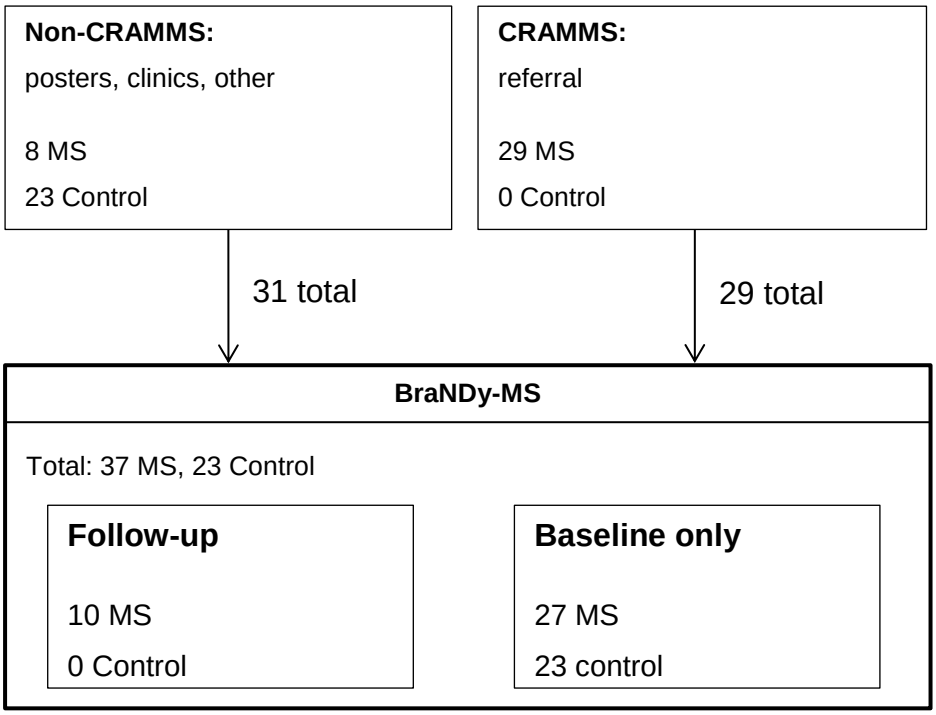


Figure 35. Flowchart of recruitment pathways.

Table 20. Demographic information for the MS group.

Variable	Total	Mean	Median	Standard Deviation	Range	Interquartile Range
N	37					
Sex	F=30, M=7					
Handedness	L=6, R=31					
Ethnicity [a]	WE=37					
MS Subtype [b]	RR=22, SP=15					
TWMLL (ml) [c]		7.77	4.81	8.87	0.19 to 32.55	1.68 to 9.48
Disease Duration (years)		17	15	10	2 to 40	9 to 25
Age at Date of First Scan (years)		48	49	11	25 to 69	38 to 57
Age of Onset (years)		31	33	6	17 to 41	29 to 36
Full-Time Post-Compulsory Education (years)		2	2	2	0 to 7	0 to 3

[a] WE = White European.
[b] RR = Relapsing-Remitting, SP = Secondary Progressive.
[c] TWMLL = Total white matter lesion load.

Table 21. Demographic information for the control subjects, n=23.

Variable	Total	Mean	Median	Standard Deviation	Range	Interquartile Range
N	23					
Sex	F=17, M=6					
Handedness	L=3, R=20					
Ethnicity [a]	WE=21, A=1, BI=1					
TWMLL (ml) [b]		0.42	0.11	0.87	0.00 to 3.93	0.00 to 0.47
Age at Date of First Scan (years)		42	42	12	26 to 63	32 to 50
Full-Time Post-Compulsory Education (years)		3	4	2	0 to 9	2 to 6

[a] WE = White European, A = Arabic, BI = British Indian.
[b]TWMLL = Total white matter lesion load.

Of the MS group, 12 (37%) were taking disease modifying drugs (3 natalizumab, 3 dimethyl fumerate, 3 fingolimod, 2 beta-interferons and 1 teriflunomide). Seventeen of the MS group (46%) were also taking antidepressants of various types (8 SSRI, 9 tricyclic, 1 SNRI, 1 atypical). Alongside these, the members of the MS group took a range of drugs for symptom relief which may also be relevant to cognition or fatigue; for example, muscle relaxants, anti-epileptics and anti-vertigo drugs, which are each sedative in their effect. Five of the control group (22%) were using antidepressants (3 SSRI, 2 tricyclic).

All volunteers completed the tests to receive an MSFC summary score. In MSFC scores, lower numbers represent more impairment and higher scores represent less impairment and better performance. All subjects received an MSFC score, even where some tests were not completed due to the way the MSFC score is calculated. The MSFC Z-scores based on the present sample

ranged from -1.34 to 1.30. Based on the Shapiro-Wilk test (p=0.262), and inspection of the histogram (figure 36A), the MSFC data followed a normal distribution. When calculating the Z-scores using the reference population, the range of the Z-scores was -2.19 to 0.34, with a mean of -0.72 and standard deviation of 0.66. These data were also normally-distributed (Shapiro-Wilk test: p=0.288; figure 36B). A one-sample T-test confirmed the difference in means (t=-6.636, p=<0.0001).

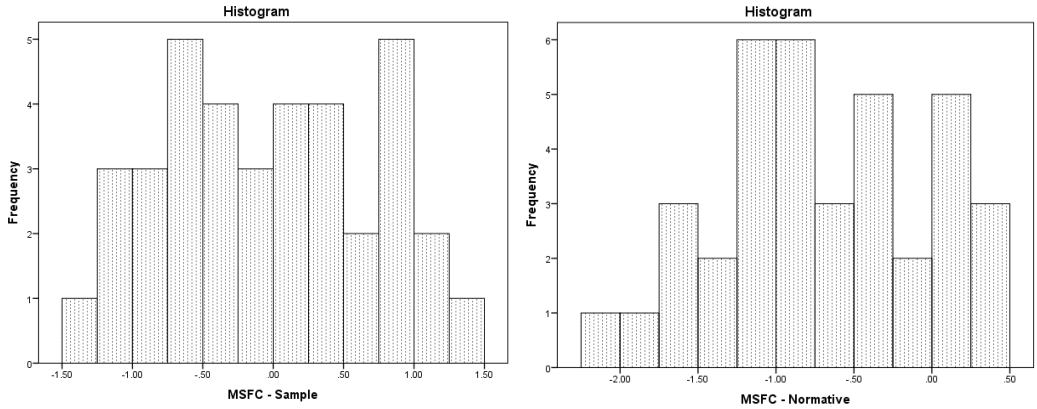


Figure 36. Histograms for the MSFC Z-scores data based on (left) the sample mean and standard deviation and (right) the normative dataset mean and standard deviation. Both follow a normal distribution.

The raw scores for the neuropsychological tests and MSFC subscores were tested for normality using the Shapiro-Wilk test and visual assessment of histograms. The only variables which were not normally-distributed were the 25FW for the MS group, the PSQI for the control group and the BDI for the control group. Statistical information for the neuropsychological test raw scores is displayed in table 22. Some subjects did not complete all neuropsychological tests due to time constraints, but in cases where tests needed to be missed, the cognitive tests were prioritised.

Table 22. Statistical information for the raw scores from the neuropsychological tests. Means and standard deviations are given because most data were normally-distributed.

	MS				Control			
	N	Mean	Standard Deviation	Range	N	Mean	Standard Deviation	Range
MSFC	37	-0.38	0.69	-1.74 to 0.83	23	0.64	0.43	-0.07 to 1.30
PASAT	32	36.31	2.44	0.00 to 59.00	23	44.87	2.43	20.00 to 60.00
9HPT	37	25.03	7.09	16.37 to 48.82	23	22.23	6.71	14.83 to 39.72
25FW	29	8.78	6.26	3.79 to 33.10	13	4.59	0.87	3.16 to 6.55
PSQI	37	8.35	3.42	2.00 to 17.00	22	5.27	4.26	0.00 to 16.00
MFIS	36	54.25	19.98	1.00 to 81.00	22	11.00	10.08	0.00 to 37.00
BDI	33	18.21	12.37	1.00 to 50.00	18	4.83	7.09	0.00 to 23.00
SDMT	33	41.91	13.09	16.00 to 68.00	23	57.00	10.74	34.00 to 74.00
NFI-MS	37	48.24	10.57	17.00 to 65.00	21	13.95	12.16	0.00 to 46.00
ANT								
Alerting	34	-0.97	14.53	-34.00 to 32.00	21	17.33	18.57	-20.00 to 50.00
Orienting	34	20.58	24.19	-16.00 to 101.00	21	36.29	23.95	2.00 to 81.00
Conflicting	34	99.12	49.05	41.00 to 211.00	21	170.90	72.45	48.00 to 272.00

Of the 33 MS participants who completed the BDI, 9 (27%) fell within the non-depressed range, 8 (24%) indicated mild depression, 10 (30%) indicated moderate depression, and 6 (18%) indicated severe depression. All but 2 of the 18 controls who completed the test had scores in the non-depressed range (88%). The two controls in the depressed range scored 23, indicating low-moderate depression.

Of the 37 MS participants completing the PSQI, 10 (27%) were classed as not having poor sleep, with a score of less than 5, and 27 were classed as having poor sleep (score >5; 82%) [317]. Of the 22 controls completing the test, 15 (68%) did not suffer poor sleep and 7 (32%) did.

For the MS group data, these raw scores for each test were then converted to Z-scores based on normative data and one-sample T-tests used to test

equality of the means to 0 (table 23). The sources for normative data were chosen only based on correspondence to the present demographic (mostly middle-aged women with RRMS) from the following sources: MSFC and subcomponents [311], PSQI [319], MFIS [327], BDI [328], SDMT [329]. Z-scores for all variables had normal distributions, except for the 25FW. That the MSFC, PASAT and 9HPT T-tests were significant while the 25FW was not suggests that the lower MSFC scores (representing greater impairment) were driven primarily by the greater-than-average impairment in PASAT and 9HPT tests, while 25FW scores were generally similar to less cognitively impaired of the MS group. The PSQI and BDI were not significantly different from the reference population mean, which is expected because they do not focus on any cognitive component. SDMT is a test of cognition and the sample had significantly poorer performance. For the MFIS, the sample scored more highly, corresponding to a greater level of fatigue, but this result was barely significant.

Table 23. Statistical information for the Z- scores from the neuropsychological tests. Means and standard deviations are given because most data were normally-distributed.

	MS				
	N	Mean	Standard Deviation	Range	One-Sample T-Test, p
MSFC	37	-0.72	0.66	-2.19 to 0.34	<0.001
PASAT	32	-0.72	1.14	-3.73 to 1.16	0.001
9HPT	37	-0.58	1.13	-4.22 to 1.7	0.003
25FW	29	-0.06	0.55	-0.50 to 2.07	n/s
PSQI	37	-0.05	0.74	-1.43 to 1.83	n/s
MFIS	36	0.47	1.19	-2.43 to 2.02	0.021
BDI	33	-0.25	1.18	-1.88 to 2.78	n/s
SDMT	33	-0.67	1.06	-2.78 to 1.45	0.001
NFI-MS [a]	37				
ANT [a]	34				

[a] Appropriate normative data from a comparable sample were not available.

Basic group comparisons in demographic and neuropsychological data were performed using one-way ANOVA (table 24). All test scores were significantly different between groups, and in the expected direction (of greater impairment in MS). Both the age and education variables were not significantly different between groups, suggesting that the groups were reasonably matched in these aspects. Figure 37 shows boxplots for the age and education comparisons.

Table 24. Group comparisons in demographic information and neuropsychological test scores.

Variable	Degrees of Freedom (Between, Within)	F Ratio	p
MSFC	1, 58	39.72	<0.001
PASAT	1, 53	5.82	0.02
9HPT	1, 58	38.24	<0.001
25FW	1, 40	18.24	<0.001
PSQI	1, 57	9.31	0.003
MFIS	1, 56	88.84	<0.001
BDI	1, 49	17.77	<0.001
SDMT	1, 54	20.77	<0.001
NFI	1, 56	126.46	<0.001
ANT			
Alerting	1, 52	16.38	<0.001
Orienting	1, 52	5.46	0.023
Conflicting	1, 52	18,90	<0.001
Age at Date of First Scan	1, 58	3.69	n/s
Years of Full-Time Post-Compulsory Education	1, 58	3.14	n/s

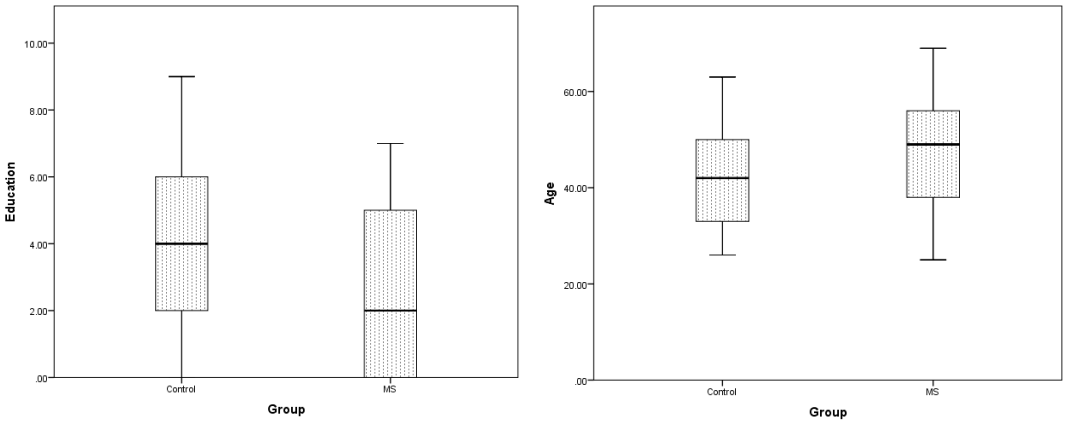


Figure 37. Boxplots showing no significant group differences in education (years of full-time post compulsory education) and age (years).

Due to the high depression and fatigue scores in some participants in the MS group, it was pertinent to investigate the confounding effects of these variables on the main cognitive scores, as well as possible effects of depression on the fatigue scores. All variables tested for correlation were normally distributed. The PASAT and SDMT were uncorrelated with either the MFIS, NFI or BDI in both uncorrected measures (Pearson correlation) and in measures corrected for age, gender and education (partial correlation). This suggests that there is no confounding effect for the two most important cognitive scores. However, the summary MSFC score correlated negatively with the MFIS ($r=-0.56$, $p=<0.001$) and negatively with the NFI ($r=-0.35$, $p=0.035$) fatigue scores in uncorrected measures (in both cases, greater functional impairment correlated with greater fatigue). However, this effect was not present in corrected correlations, where the MSFC no longer correlated with the MFIS ($r=-0.31$, $p=0.21$) or NFI ($r=0.07$, $p=0.78$). The MSFC weakly correlated negatively with the BDI in age- and education-corrected

partial correlations (less impaired people with MS tended to have less depression; r=-0.49 p=0.04). This suggested that results involving the MSFC score may be partially explained by differences in depression or fatigue scores and should be interpreted with that in mind. These relationships are shown in figure 38. A third potential confound, sleep quality [318, 319], was significantly correlated with the NFI in the expected direction but only in the uncorrected correlations (r=0.41, p=0.01).

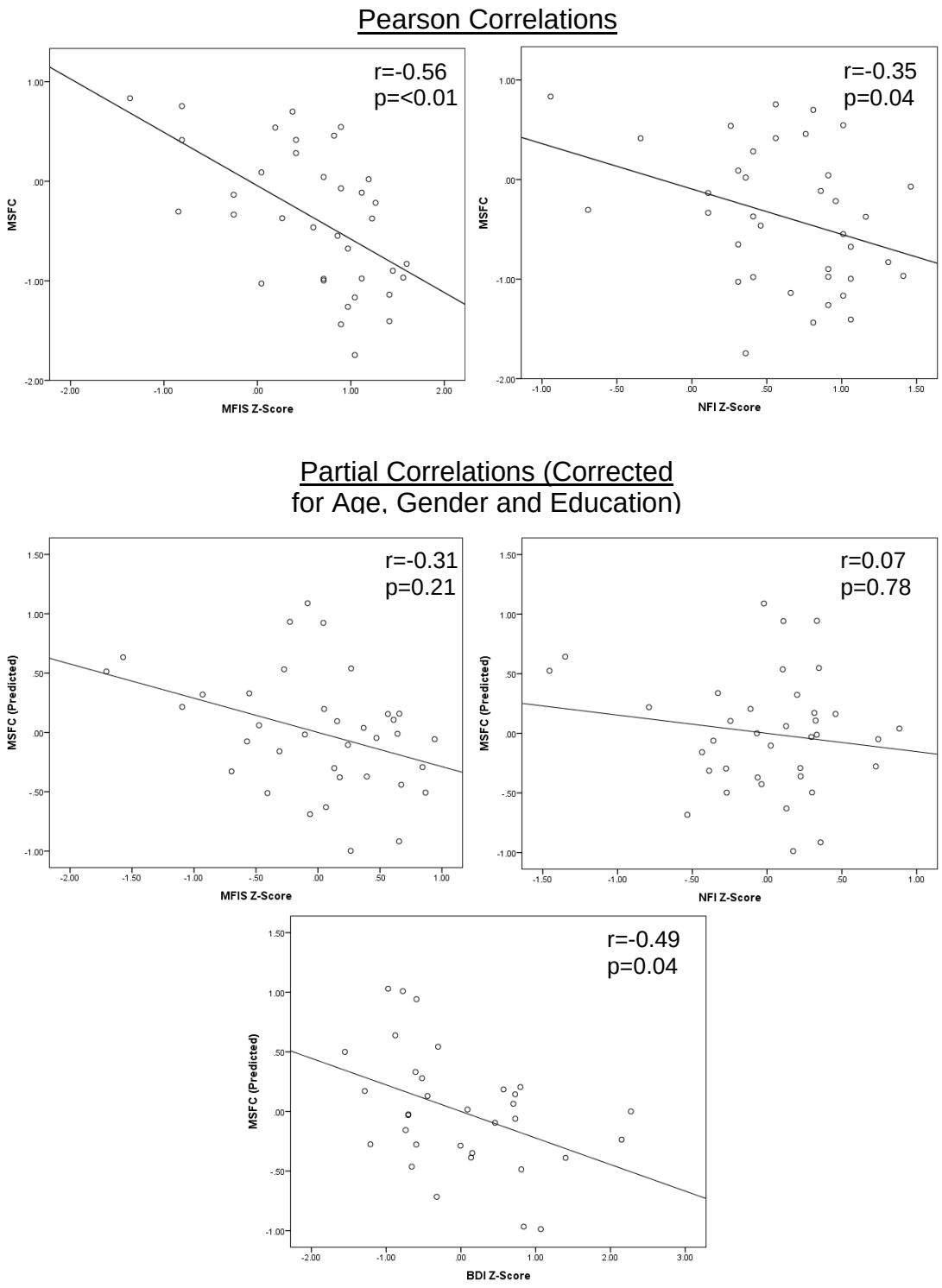


Figure 38. Corrected (partial) and uncorrected correlations between MSFC and fatigue and depression scores in the MS group only. Higher MSFC score indicate less impairment/better performance, whereas higher scores in the BDI, NFI or MFIS indicate more impairment/worse performance. The scatter plots demonstrate that, while the distributions of MSFC and BDI

scores were fairly uniform and had no major outliers, the NFI and MFIS questionnaires produced several outliers in the MS group with low fatigue scores. However, these data points were unlikely to have driven the significant uncorrected correlations as they appear close to the line of best fit and so probably did not cause a false line of regression. They were also not large enough outliers for significance to be retained in the corresponding partial correlations.

Summary

The results presented in this section provide an understanding of the composition of the participant groups and the extent of functional impairment. The main finding was that the studied cohort comprises people with MS with a range of cognitive abilities from very mild to severe, and, as the comparisons to normative datasets show, the cohort is generally more cognitively impaired by approximately 1 standard deviation than normative populations for fatigue, cognitive and some motor functions, although some subjects had Z-scores for the PASAT and SDMT more than 2 standard deviations from the mean. While some subjects had high depression scores, the main cognitive tests were uncorrelated with suspected confounds like fatigue and depression, which means that the end results of my experiments are unlikely to be driven by these.

The goal was to evaluate the suitability of the studied groups for the purpose of answering the main thesis research questions. While the above factors, particularly the higher frequency of cognitive impairment in the group, may limit the generalisability of the findings to cognitively preserved individuals with MS, they may also allow us to study cognition with greater accuracy due to the range of abilities.

The lower mean of the Z-scores based on normative data suggests that the subjects have a greater level of impairment than the general population of people with MS. This was expected due to the screening process for volunteers recruited through the CRAMMS trial being selective toward those with cognitive impairment. It may mean that my results are more applicable to people with MS with greater levels of functional impairment than to the average person with MS.

Non-normality of the PSQI and BDI scores in the control group was expected because controls will tend to often score 0 or very low on these tests, both of which are bounded at 0. The reason for the right-skewed distribution in the 25FW for the MS group is not as clear, but may be due to chance or to a possible tendency for people with MS to tire and hence perform the second repetition more slowly.

One potential limitation to the neuropsychological testing protocol was that, due to both its length (approximately 1 hour) and the propensity for fatigue in the MS group participants, there was potential for subjects to tire during the course of the testing. To alleviate this, the tests were administered in order of importance, with cognitive tests first and motor tests performed last. Some subjects also reported feeling tired after the MRI scan, which may have degraded their performance on some tests. While the effect of this intra-session fatigue is unlikely to have a large impact within the group, it may have negatively impacted the comparisons to normative datasets, and may have been one factor driving the relatively poor cognitive performance of the cohort.

One limitation of the fMRI protocol I carried-out was that, during eyes-closed fMRI sequences there may be an increased tendency for people to fall asleep, which may be further exacerbated in people with fatigue. Resting-state neural signals measured by fMRI are altered during sleep [330], and so this may have affected my findings. An eyes-closed protocol was followed so that the resting-state fMRI data did not include a visual processing component. This may have been somewhat mitigated by the communication with the volunteer during scanning in between sequences through a speaker inside the scanner. Others have attempted to mitigate this effect by having volunteers fixate on a cross central in their visual field [331].

7.3. Methods

All experiments in this chapter used fMRI data from the BraNDy cohort described in section 7.2. All subjects had complete data and none were excluded based on incomplete or corrupted MRI. The scores from the PASAT, SDMT, ANT, MFIS and NFI tests also described in section 7.2 were used.

Data Quality

It was important that I ascertain the quality of the MRI data before conducting my analyses. Errors in the data introduced during analysis can propagate to produce spurious results. To address this problem, myself and Dr Xing Yue developed and applied a script for fMRI QA which incorporated data quality best practices from other studies including the fBIRN initiative [332]. The script was written in bash and MATLAB and produced the following outputs:

- Frequency spectrum.
- Mean image.

- Standard deviation image.
- An even-odd difference image.
- Signal-to-fluctuation noise ratio image.
- Mean timecourse.
- Signal-to-noise ratio.
- A text file of motion parameters in each of the three axes for translation and rotation.
- The results of an ICA using FSL MELODIC.

The script was applied to the fMRI data for each subject and then each of the outputs were viewed to identify any datasets with excessive noise or motion. This was based on a judgement by myself, mostly from looking at the frequency spectrum and mean timecourse, and confirmed by the other outputs. The following criteria were applied to improve the robustness of my subjective assessment: (1) “is most of the power in the spectrum below 0.1 Hz?” (2) “are there few other large peaks in the spectrum indicating sources of noise, such as at the typical frequencies for respiration or heart rate?” (3) “are there no large repetitive deviations in the mean timecourse indicating motion or other sources of noise?” (4) “does the mean image appear to show clearly structures in the brain such as the ventricles and a contrast between white and grey matter?” (5) “is the temporal fluctuation image relatively dark?” (6) “are there no structures visible in the static spatial noise image, for example the skull outline?” (7) “do any other outputs appear out of the ordinary?”. No quantitative thresholds could be used, firstly because the judgement involves looking for structure in images, which is difficult to automate and secondly because some compromise can be made in cases where some outputs

appear “good” but others do not. There is room for human error in this approach, although in my experience it is usually easy to identify noisy datasets using this method, as shown by the examples given in section 7.4.

Image Processing

The steps for segmentation using the HOA and fMRI preprocessing described in section 7.2 were applied. To produce 126x126 adjacency matrices for each subject, the steps described in section 7.2 were applied.

Graph Analysis

Before performing the main analyses it is useful to verify that the thresholded graphs follow the defining properties of small-world networks: a low characteristic path length, a high clustering coefficient and the presence of high-degree hubs [333]. If the generated networks do not have these properties, they are unlikely to be reflecting true biological structures. To investigate this, 1000 random and one lattice graph of the same average size and order as the brain MRI graphs for MS participants were first generated at each density using MATLAB (with the BCT methods for randomisation). The same integration procedure as for the human MRI data was then applied to ascertain a summary score for the random and lattice graphs.

The two types of association matrix contained all the information necessary to describe weighted graphs, where each ROI is modelled by a node and each connectivity value is modelled by a weighted edge. To the functional connectivity matrices, I applied Fisher’s Z transformation (since the r-values are non-linear), and to the SD matrices I re-scaled the weights such that they ranged between 0 and 1. To remove spurious or weak connections and retain

the strong ones, I applied a minimum density threshold to each graph. The density of a graph is the ratio of the number of existing edges to the number of possible edges. In order to avoid arbitrarily selecting a single threshold and to improve the reliability of my results, a range of density thresholds was applied. The range of thresholds used was 0.2 to 0.5, in increments of 0.02, giving 15 total densities. This range was chosen with the aim of producing results which are reliable and are not based on graphs which are (a) fragmented and (b) too dense to have small-world properties [334, 335]. The integral of each metric score across this range (as in [336]) produced a summary score which was less susceptible to large possible variations in reliability with density, as concluded from chapter 6. A MATLAB script was written to implement the density threshold by applying a binary search algorithm (one which repeatedly targets the centre of the data to divide the search space in half) on the range of raw thresholds (in the range 0-1) to find the threshold which produced a graph with a density as close as possible to the target density. Once a graph had been appropriately thresholded, it was used to calculate a set of global metrics using the Brain Connectivity Toolbox [337] (BCT; a library of MATLAB functions for manipulating graphs). The chosen metrics were selected based on their reproducibility and frequency of use in previous literature [338].

1. Clustering coefficient.
2. Characteristic path length.
3. Global efficiency.
4. Small-worldness.
5. Modularity.

The above metrics were also calculated for 1000 randomly-generated graphs which preserved the average degree sequence of the combined sample. To do this, an Erdős–Rényi graph was created for each subject of the same size and degree sequence on an individual basis. This was repeated until 1000 random graphs had been created. The process of generating these graphs was done with a custom MATLAB script and the BCT.

Statistical Analyses

There were three main outcomes: the group difference in graph metrics, the relationships between graph metrics and cognition / fatigue scores and the analysis of those relationships in individual regions. All variables were first tested for normality using Kolmogorov-Smirnov tests and by visual inspection of histograms. For the analysis of group differences in summary graph metrics, the normalised metric scores were entered into multiple one-way ANCOVAs with group as a factor and age, sex and education as covariates. For the figures, metric scores were transformed to Z-scores so that they could be shown on the same scale.

To test relationships between graph metrics and neuropsychological measures in the MS group, I first checked for normality and for linear relationships between the variables of interest using scatter plots and tests for correlation. I then performed linear regression analyses with the neuropsychological measures (PASAT, SDMT and ANT subscores) or fatigue measures (MFIS and NFI-MS) as the dependent variables and the network metrics entered as the independents. The “enter” method was used because I didn’t know the relative importance of the independent variables. The level of statistical significance for all tests was set at 5%.

I performed a second set of regression analyses which were the same but with the covariates of no interest (age, gender, education, PSQI and BDI) entered as the first block of independents and the graph metrics entered as a second block. This was done separately from the no-covariates tests because there is a potential for overfitting when the number of subjects per independent variable is low (a common rule-of-thumb is 10 per variable) and therefore a likelihood of overestimating the model R^2 in these tests [339].

7.4. Results

Data Quality

The mean relative displacement (between successive volumes) was 0.125 mm (SD 0.051) for controls and 0.203 mm (SD 0.344) for the MS group. Figure 39 shows an example of the outputs of the QA script from a “good” dataset before any preprocessing had taken place. In contrast, figure 40 shows an example of a “bad” dataset, also prior to preprocessing. Based on inspection of these outputs two datasets were excluded from further analysis due to excessive motion during the fMRI acquisition. Otherwise, the quality of the data was good.

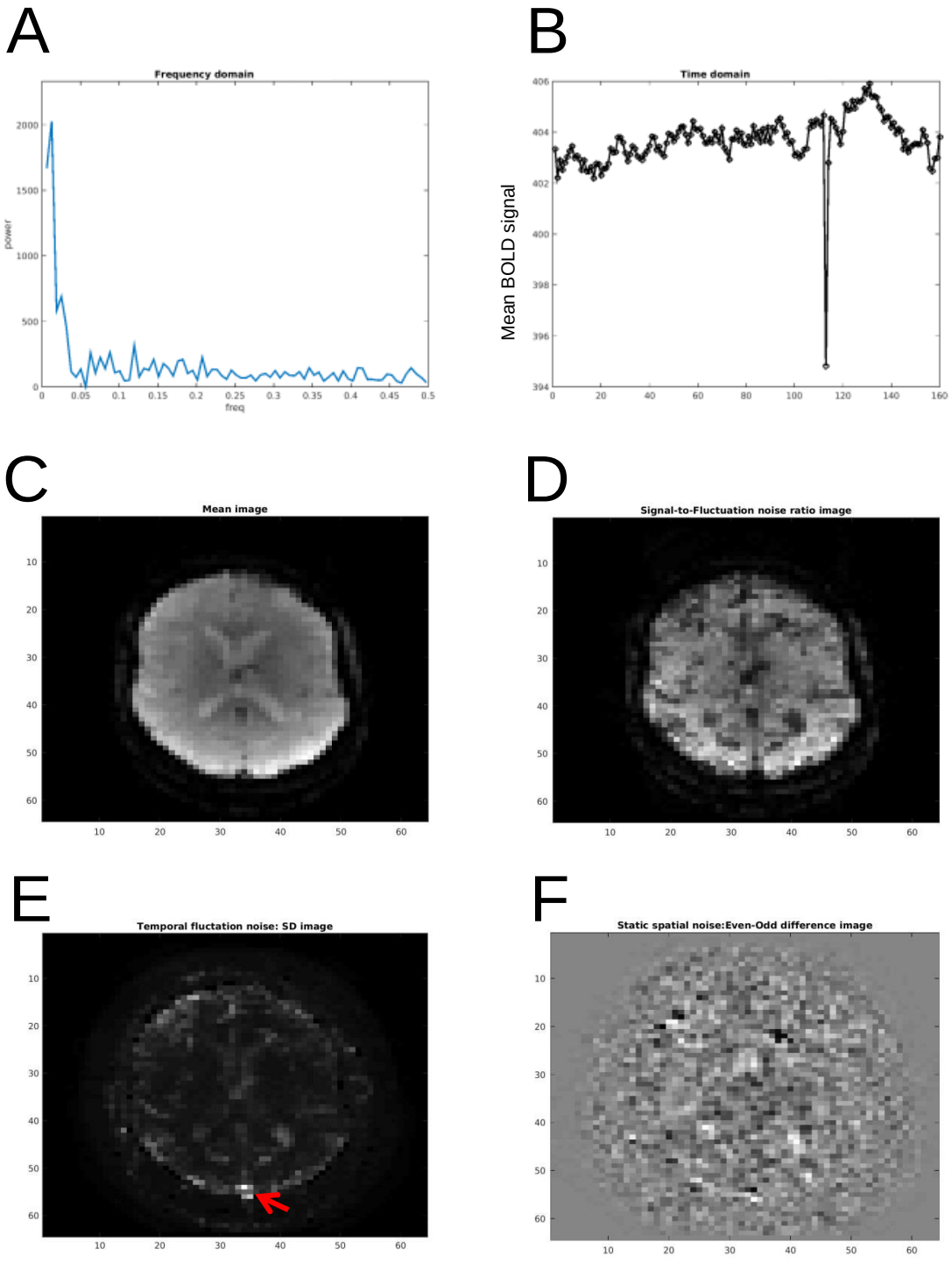


Figure 39. Outputs of the script used for fMRI quality assurance which were checked for each dataset. The figure shows an example output from a good dataset. (A) Frequency spectrum. (B) Mean timecourse. (C) Mean image. (D) SFNR image. (E) SD image. (F) Even-odd difference image. This was deemed an example of a good dataset because the majority of the power spectrum lies below 0.1 Hz, the known range for spontaneous BOLD fluctuations [340].

The clear dip in the mean timecourse in (B) is a motion artefact, as confirmed by manual inspection of the relevant volumes. In volumes such as this, the motion would be non-destructively scrubbed by effectively ignoring it in the nuisance regression. The mean image shows, by not being blurred, that the amount of motion during the scan was minimal. For the temporal fluctuation (SD) image, the darker it is, the better, as high intensity voxels represent high standard deviations over the timecourse. This explains why the superior sagittal sinus (red arrow) appears bright in this image. In the ideal even-odd difference image, no structure should be visible because it represents only static spatial noise.

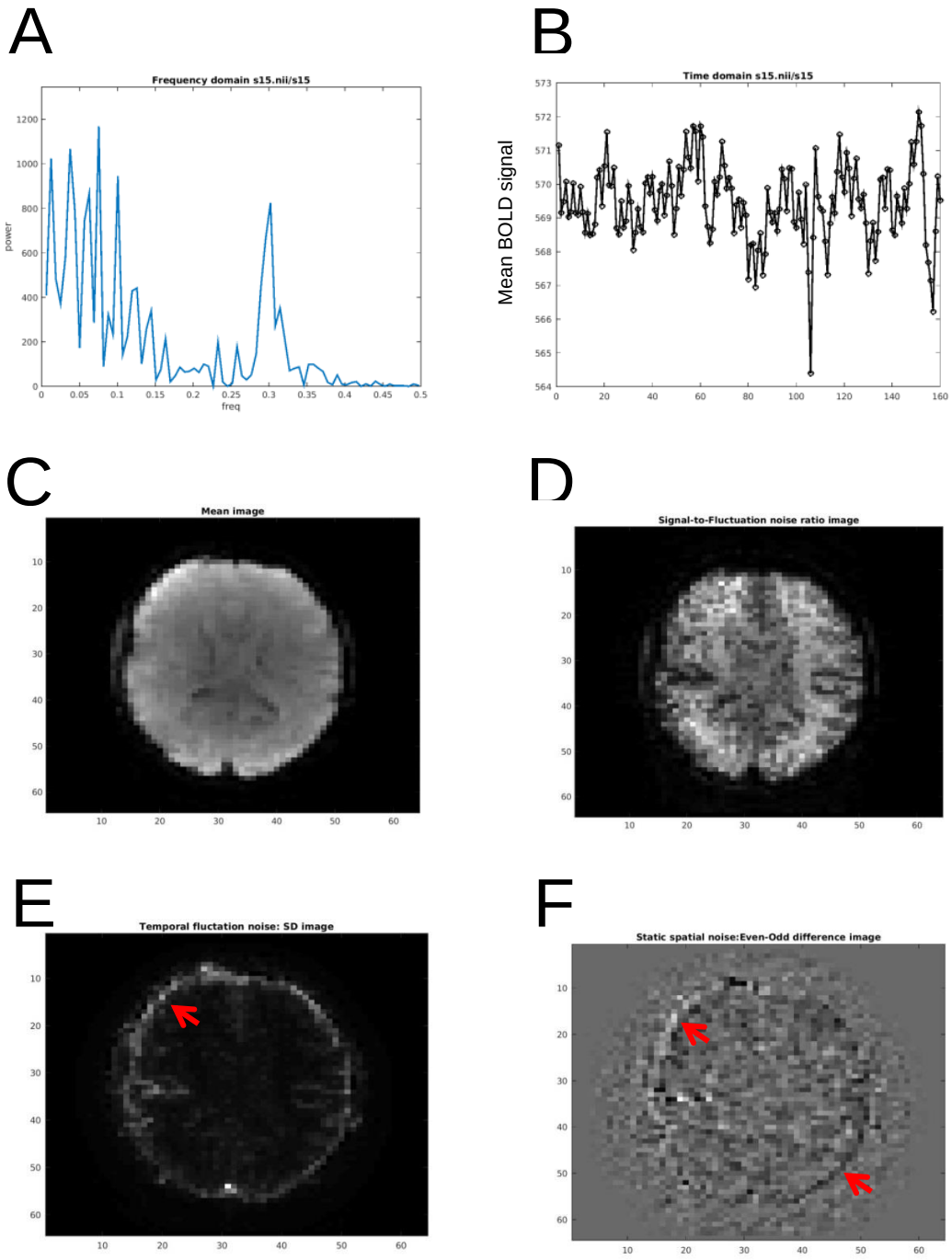


Figure 40. Example of a bad dataset, which would be excluded from the analyses. The indications that this is a noisy dataset are, firstly, that the frequency spectrum includes a large peak at 0.3 Hz, which may be due to respiration. Second, the mean timecourse, although the scale is smaller than in figure 39, does show larger variations in signal. Third, there is a clear ring in the SD image which may be related to blood flow or motion. Finally, the outline of the skull can be seen in (F), suggesting consistent motion.

Small-World Structure

The clustering and path length for the random, lattice and brain MRI graphs were compared in figure 41A and 41B. Figure 41C shows histograms of degree distribution pooled across all control subjects for the densities 0.2 to 0.5.

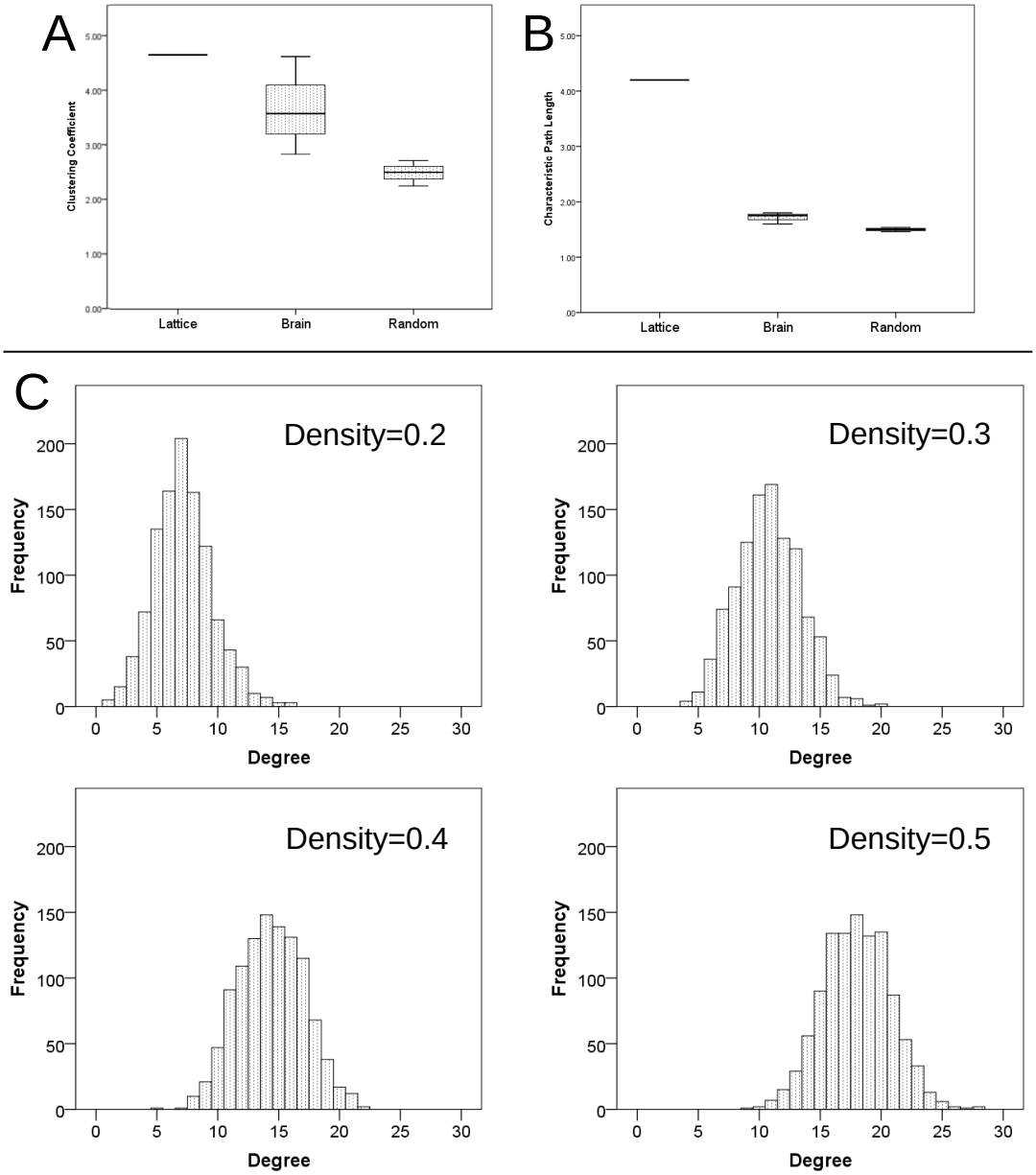


Figure 41. Confirmation of small-worldness in the brain networks. (A& B) Comparison of clustering and path length between regular (lattice), random, and brain networks. The clustering and path length metrics for brain networks fall between random and lattice graphs of the same size and order. The scores for these metrics also meet the common heuristic for small-worldness of: path length ≈ 1 and clustering > 1. (C) Histograms of nodal degree for all MS subjects for the densities 0.2, 0.3, 0.4, 0.5. This shows that degree follows a normal distribution with relatively few high-degree nodes which serve as “hubs”, and that this is consistent throughout the range of measured graph densities. As the density increases, as

well as the mean degree increasing as expected, the distribution becomes more platykurtic, meaning that more of its variance is attributable to frequent modestly-sized deviations from the mean.

Group Differences in Network Organisation

I then tested for group differences in the five network metrics. Boxplots of the normalised metric scores for each group are shown in figure 42. The ANCOVA analysis revealed that, when correcting for age, gender and education, highly significant group differences in the characteristic path length, global efficiency and modularity metrics, with the clustering coefficient barely reaching significance and the small-worldness index being nonsignificant (table 25).

Table 25. Outcome of the ANCOVA analysis testing for group differences in graph metrics with age, gender and education as covariates.

Metric	F	p
Characteristic Path Length	15.944	<0.001
Modularity	7.063	<0.001
Small-Worldness	2.670	0.056
Global Efficiency	29.289	<0.001
Clustering Coefficient	3.146	0.032

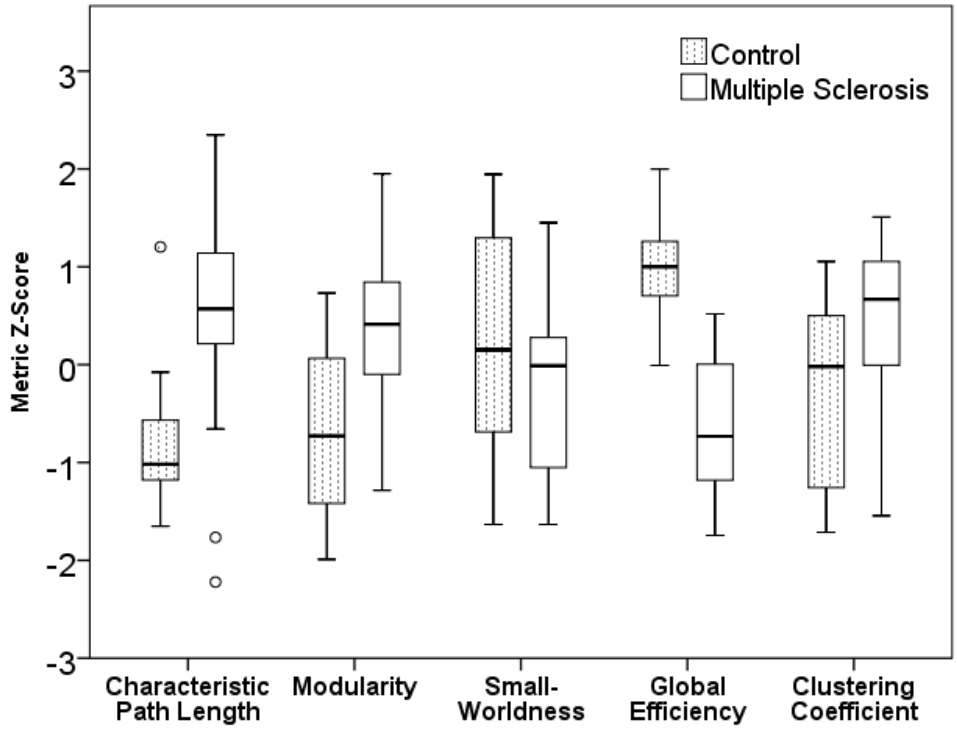


Figure 42. Boxplots for the normalised metric scores for each group (uncorrected for demographic variables).

I then tested for correlation between T2LL and the 5 measured graph metrics, with the hypothesis that one or more metrics would be correlated with lesion load, thus supporting the idea that white matter damage drives a network disconnection phenomenon. I found one significant negative correlation between modularity and T2LL (r=-0.343, p=0.038). This relationship is in the expected direction given the other findings. The correlations for path length, small-worldness, global efficiency and clustering coefficient were not significant (in order: r=0.212, p=0.207; r=-0.46, p=0.789; r=-0.53, p=0.755; r=0.172, p=0.309).

Relationships Between Network Organisation and Neuropsychological Measures

All graph metrics had approximately normal distributions and passed a Shapiro-Wilk test. The initial tests for linear relationships between the variables showed significant relationships between the cognitive test scores and most graph metrics, and between the fatigue measures predominantly with the characteristic path length (figure 43). More importantly, correlations between the independent variables to be entered in the regressions (network metrics) were negative, except for the correlation between modularity and characteristic path length ($r=0.328$, $p=0.048$), although this was one of 10 tests. Among controls, the correlations were negative except for the correlation between small-worldness and global efficiency ($r=0.420$, $p=0.046$).

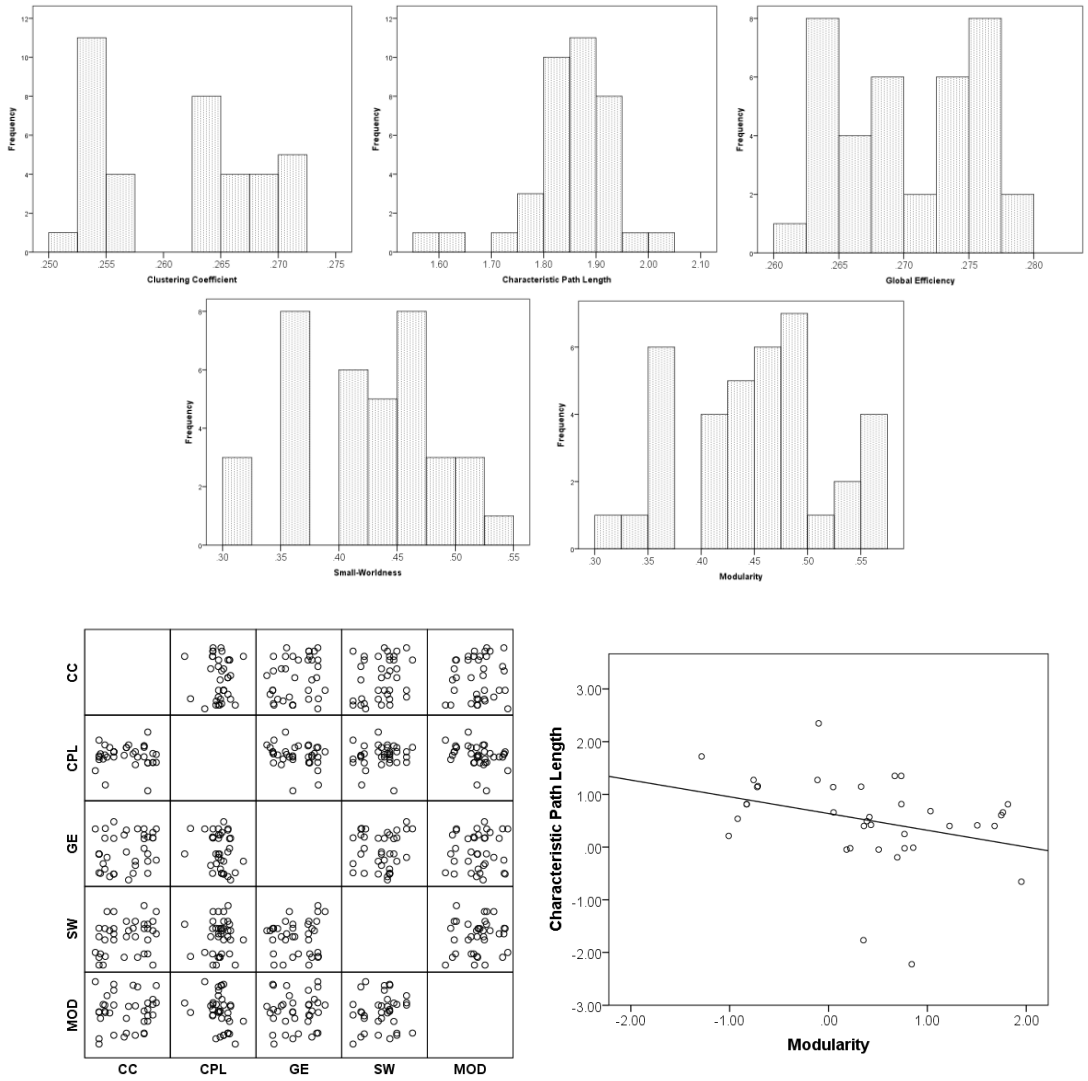


Figure 43. Histograms showing normal distributions for network metric variables in the MS group (top), scatter matrix of correlations between the independent variables entered in the regressions (bottom left) and a scatterplot of the significant correlation between modularity and characteristic path length (bottom right).

With the linear regression analyses, I found that the final models predicted the PASAT and SDMT cognitive tests to a statistically significant degree, but not the ANT subscores (alerting, orienting and conflicting; table 26) or fatigue measures. The variables which significantly contributed to predicting 55.6% of the variance for the PASAT were small-worldness, modularity, characteristic path length and clustering coefficient (each having $p<0.05$ in a one-sample t-

test on the unstandardised coefficients). The direction of the relationships indicated that individuals with greater levels of impairment on the PASAT had increased clustering and modularity, longer average path lengths and less small-worldness. The regression coefficients were as follows: path length -0.261, modularity -0.074, small-worldness 0.050, global efficiency 0.457, clustering 0,278.

The variables which significantly contributed to predicting 60.0% of the variance for the SDMT were small-worldness, global efficiency and characteristic path length. The direction of the relationships indicated that individuals with greater levels of impairment on the SDMT had less network small-worldness, lower global efficiency and longer average path lengths. The regression coefficients were as follows: path length 0.575, modularity 0.296, small-worldness 0.495, global efficiency 0.423, clustering 0.175.

Table 26. Outcomes from the multiple linear regressions for cognitive and fatigue measures in people with MS (final models) not including covariates. All independents were entered into the model.

	Network Metrics Model		
Cognitive / Fatigue test, dependent	F	P	R^2
PASAT	8.752	<0.001	0.556
SDMT	16.384	<0.001	0.600
ANT Alerting	1.675	0.175	0.237
ANT Orienting	0.583	0.713	0.097
ANT Conflicting	0.760	0.586	0.123
MFIS	0.587	0.710	0.089
NFI-MS	0.544	0.742	0.081

In the tests including the covariates of no interest, the R^2 estimates were much higher, but were likely overestimated (table 27) [339]. The stronger

predictions for fatigue measures suggest that the demographic variables were much better predictors than the graph metrics. In controls, the linear regressions showed no significant predictions for any of the independents (table 28).

Table 27. Outcomes from the multiple linear regressions for cognitive and fatigue measures in people with MS (final models) including covariates. All independents were entered into the model.

	First model (covariates of no interest only)			Second model (network metrics and covariates of no interest combined)		
Cognitive / Fatigue test, dependent	F	P	R^2	F	P	R^2
PASAT	0.967	0.458	0.168	6.452	<0.001	0.773
SDMT	0.451	0.808	0.089	4.721	0.002	0.724
ANT Alerting	0.597	0.702	0.115	2.830	0.022	0.616
ANT Orienting	1.186	0.347	0.205	1.715	0.071	0.484
ANT Conflicting	0.711	0.621	0.328	1.880	0.067	0.528
MFIS	11.716	<0.001	0.693	6.259	<0.001	0.629
NFI-MS	4.349	0.005	0.446	2.989	0.015	0.576

Table 28. Outcomes from the multiple linear regressions for cognitive and fatigue measures in controls (final models) including covariates.

	First model (covariates of no interest only)			Second model (network metrics and covariates of no interest combined)		
Cognitive / Fatigue test, dependent	F	P	R^2	F	P	R^2
PASAT	1.116	0.402	0.317	2.118	0.166	0.752
SDMT	1.005	0.456	0.295	2.659	0.104	0.792

ANT Alerting	1.594	0.241	0.420	1.849	0.233	0.755
ANT Orienting	0.808	0.577	0.269	0.980	0.535	0.620
ANT Conflicting	0.850	0.543	0.589	0.860	0.604	0.589
MFIS	0.502	0.769	0.173	0.533	0.823	0.432
NFI-MS	0.499	0.771	0.185	0.450	0.873	0.429

7.5. Discussion

In this chapter I aimed to investigate differences in graph metrics between MS and control groups, and how they relate to measures of cognitive impairment and fatigue. I began by testing the quality of the fMRI data and found that they were sufficient to reasonably eliminate poor data quality as a possible confounding effect. I then tested that the constructed graphs conformed to common definitions of small-world networks. I showed that the data from human subjects were more clustered than random graphs, but less clustered than lattice graphs, and that they had similar average path lengths to random networks, which were significantly shorter than in a lattice. Across the range of relevant densities, the degree distribution followed the expected pattern for small-world biological networks. Together, these results suggest that the data meet the requirements for small-worldness over the range of tested densities, and hence likely do measure a biological network, rather than predominantly reflecting noise. While it should be noted that small world properties have also been detected in white noise correlation matrices [341], our other findings such as observed relationships to neuropsychological variables serve to reinforce this. These findings replicate previous data from independent samples and established the validity of the data for further analysis.

The first of the main results were from the comparison of graph metrics between MS and healthy control groups. Previous studies have reported a wide range of differences, but most commonly a reduced global network efficiency in people with MS compared to controls [178, 179, 183-186], along with other changes such as reduced small-worldness [177], longer average path lengths [184, 186] and increased network modularity [187]. My data showed significant group differences in all but one metric when correcting for age, gender and education level, and the direction of the differences agree with the findings of previous studies. The uncorrected results were less pronounced (figure 42). The direction of the observed differences all suggest less efficient transfer of information around the network; these effects are assumed to also translate biologically to network disruption and impaired function. The relative contributions of the forces driving the differences are not yet known; for example, our results do not express the extent to which the differences in graph metrics are due to reorganisation as opposed to simply a net loss of connectivity in areas of focal WM damage. Disentangling the contributions of several simultaneous processes, some of which are restorative and some of which are injurious, which also may vary in their rates of action and regionally within the brain, is a difficult problem which future research should aim to address.

The second main finding of this chapter was of the relationships between graph metrics and measures of cognition and fatigue. Based on the results of previous studies [184, 186, 187, 190, 192, 195], I had hypothesised increased network segregation and decreased network integration in individuals in the MS group with more severe cognitive deficits. In the significant models for

PASAT and SDMT, the metrics and their direction of association aligned with this hypothesis. For the PASAT, the best predictor was the clustering coefficient. A more highly clustered brain network indicates that nodes' neighbours become more closely connected within themselves. This may be the brain's response to the interruption of signals in white matter tracts by focal lesions: as long-range connections or connections between subnetworks become disrupted, networks instead become more tightly connected to their strongest remaining neighbours. Future clinical studies using the PASAT may therefore focus on this metric. Looking at the significant models for cognitive tests together, the small-worldness and characteristic path length metrics were strong predictors in both models, and therefore these metrics may be useful in studies of cognition in general outside of the PASAT. This evidence is in support of the use of network metrics in clinical studies; for example, in monitoring response to cognitive therapies.

The SDMT is a simpler task than the PASAT, which focuses more on visual scanning and attention than on the actual processing of information. The observed predictors were mostly the same as for the PASAT, with the addition of global efficiency and the absence of the clustering coefficient and modularity metrics. These differences may therefore be related to the type of task that was performed. This may suggest that variance in the clustering and modularity metrics (representing increased network segregation) being related to cognitive performance are specific to tasks which involve fast processing of information and production of speech. Also of note is that both the PASAT and SDMT were predicted strongly by education level, whilst the ANT alerting score was not. This may be explained again by the simplicity of the task,

which is highly dependent on reaction time, especially considering research which shows education has less of an impact on reaction times in simpler tasks, and more of an impact for complex tasks [342].

There were also differences between the regression results for fatigue variables compared to cognitive variables. BDI was a significant predictor for both the MFIS and NFI fatigue scores, while it was not for any of the cognitive measures. The MFIS was predicted by the modularity metric, but the NFI was not predicted by any graph metric. This may suggest that changes in network organisation measures are more relevant to cognitive impairment than to fatigue. This may be due to the physical component of fatigue and, as shown in chapter 5, cerebral WM damage has a greater relevance for cognitive impairment than for physical impairment.

The less significant or negative findings in the control group may be partly explained by the much smaller variance in neuropsychological measures and the tendency for low fatigue, depression and sleep disturbance scores. It also may be that the controls' cognitive performance was determined far less by network organisation than by other factors. It is less clear why the ANT subscores for orienting and conflicting were not as well explained by brain network metrics, especially considering that other cognitive measures were.

There may be a limitation to using brain volume as opposed to intra-cranial volume for correcting ROI volumes. The intra-cranial volume correction would have accounted for an atrophy effect, which could be present in the MS group. The brain volumes were used as they were already available as outputs of the processing steps taken.

In conclusion, I was able to successfully apply the graph-theoretic model of brain connectivity to test group differences in brain network metrics and found that all but one of the tested metrics were significantly different between groups and were related to cognitive measures. The evidence presented here supports the validity of graph metrics as markers of cognitive impairment in MS, but not as good markers of fatigue symptoms. The network changes suggest a change in network structure toward being more segregated and less integrated.

8. Responsiveness of Brain Network Properties

8.1. Introduction

There are many examples of studies showing differences in functional brain network metrics between MS and control groups at rest, as reviewed in section 4.3. However, no previous study has attempted to identify dynamic topological network changes that occur during task performance in people with MS, or investigated differences in network structure between any active task performance and rest in MS, as compared to controls. Not only might this information cast new light on the mechanisms underlying cognitive symptoms in MS, but it would provide evidence in support of the position that summary organisational measures derived from brain graphs are responsive to changes in cognitive state in the immediate term. It may also identify new imaging metrics related to cognition based on the temporal network dynamics; for example “is the way in which the switching between states takes place related to any cognitive measure?”. This corresponds to the responsiveness criterion of research question 3.

Analysis of network dynamics using the graph theoretic method is a relatively new technique. The general problem faced is that, usually, the whole BOLD timecourse is used to generate measures of connectivity from which the graph is constructed but, in order to observe changes in graph metrics during the timecourse, multiple measurements must be made from within it. The most

common way to approach this is to choose a fixed window length and to "slide" this window along the timecourse to create multiple graphs [343]. The problem with this is that the measures of connectivity used for each graph will be based on much fewer data points. Thus, the question of "what is the shortest window length from which reliable measures of connectivity can be derived?" is of importance. Previous studies have used window sizes of 30-60 seconds [344-346].

Hypotheses and Aims

- Reaction times in people with MS will decrease at a lesser rate than those of controls and may increase toward the end of the task due to cognitive fatigue effects; i.e. the learning effect will be suppressed.
- Dynamic changes in functional brain network structure during the performance of a sustained attention task will be measurable in people with MS using the graph-theoretic approach.
- In people with MS, periodic reaction times to the task stimulus will be correlated with metrics describing brain network organisation measured for the corresponding time window.
- Patients with more severe cognitive impairment, specifically in measures of sustained attention, will have the most pronounced differences in brain network dynamics from controls.

The broad aim of the work in this chapter was to determine the responsiveness of brain graph metrics to induced changes in cognitive state e.g. during sustained attention or during the transition from active task performance to resting state. This would provide evidence toward the

requirement for responsiveness described in research question 3. A secondary aim was to explore the network changes that occur during task performance and interpret them with the aim of understanding mechanisms of fatigue and cognitive impairment.

8.2. Methods

The acquisition of the task and MRI data used in this experiment is described in section 7.2. In short, brain images from healthy and MS groups were segmented using the HOA and fMRI timeseries for these regions extracted. The fMRI data used were acquired during performance of the mSDMT task. Task reaction times, measured as the interval from when the stimulus is shown to the first button press, were recorded in E-Prime and tabulated. Reaction times from trials where an incorrect answer was given were ignored.

A “window” of size 110 seconds (50 TRs), corresponding to the lowest frequency of interest [344], was created (figure 44). The window was shifted 4.4 seconds (2 TRs) repeatedly along the timeseries, resulting in 84 windows. The correlation coefficients between each pair of timeseries in each window were calculated and transformed into z-scores to create weighted undirected association matrices for each subject. From each matrix, three summary graph metrics (characteristic path length, mean clustering coefficient and small-worldness) were calculated at a range of densities and summarised by the integral of the results over the range, as in Hosseini, Hoeft [336]. These metrics were chosen on the basis that they are reproducible and frequently used throughout the literature [338]. As a benchmark, the process was

repeated for a matrix made from the average of 1000 randomly-generated networks with the same degree sequence as the test subjects.

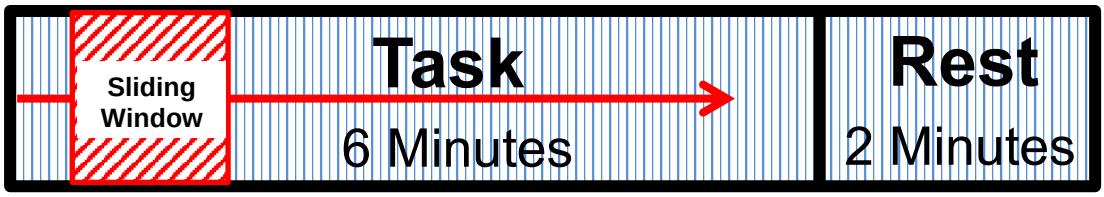


Figure 44. The fMRI task lasted 6 minutes and was followed by 2 minutes of rest. In red is the sliding window used to divide the timeseries.

To investigate how brain network dynamics are related to neuropsychological measures of attention network functioning, the difference between the mean mSDMT reaction time for the final 2-minutes of the task and 2-minute rest for each subject was tested for correlation with the three ANT subscores (alerting, orienting and executive control). The same was performed for the first 2 minutes of the task and the final 2 minutes of the task (table 31).

8.3. Results

36 of the 37 MS participants had complete task and fMRI data. The one subject without it was uncomfortable during the scan and asked for it to be skipped. All 23 control subjects had complete data. The extent of head motion was not significantly different between groups (mean relative displacement: MS mean: 0.091mm, SD 0.036; control mean: 0.967 mm, SD 0.034; two-sample t-test: t=0.587, p=0.559).

First, I wanted to confirm that reaction times did indeed change during task performance. I expected that any changes would be reflecting two conflicting processes: a learning effect which would act to reduce reaction times as

subjects became accustomed to the task and a cognitive fatigue effect which would act to increase reaction times as subjects became fatigued. I expected that the former effect would be present in both MS and control groups, but that the latter would only occur in the MS group (or occur in both, but to a greater extent in the MS group). The reaction time data are presented in figure 45. They show a learning effect for both groups but that reaction times of the MS group were consistently slower than those of controls (control mean: 2240ms, SD 837; MS mean: 2572ms, SD 999; 2-sample t-test across all trials: t=-6.522, p=<0.001).

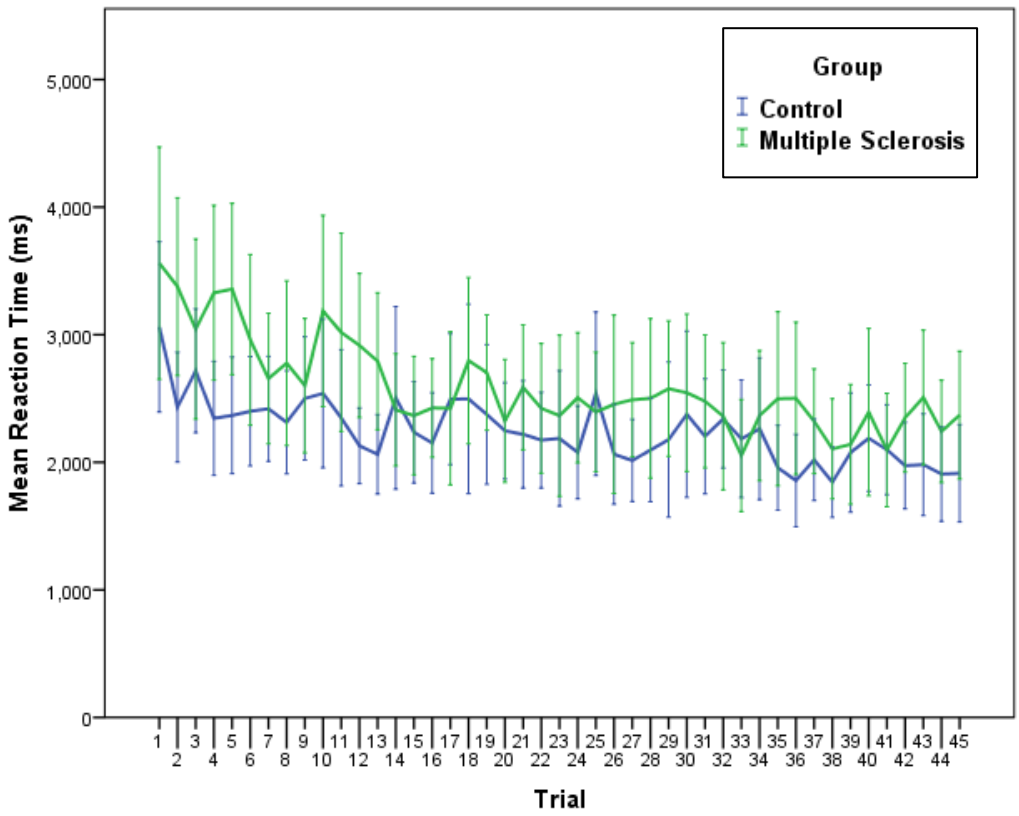


Figure 45. Reaction times for the two groups during task performance.

I then wanted to measure graph metrics in a sliding window over the course of the task and test whether or not they were related to reaction time. Z-scores

for the three metrics are plotted against time for some example subjects in figure 46. The mean Z-score across all subjects is plotted against time in figure 47. Sliding window graph metric data (minus the rest part) were resampled using MATLAB to match the sample rate of the reaction time data (45 samples for the 6-minute task period), so that linear correlations could be performed. The resampling function applied a lowpass anti-aliasing FIR filter and compensated for its delay.

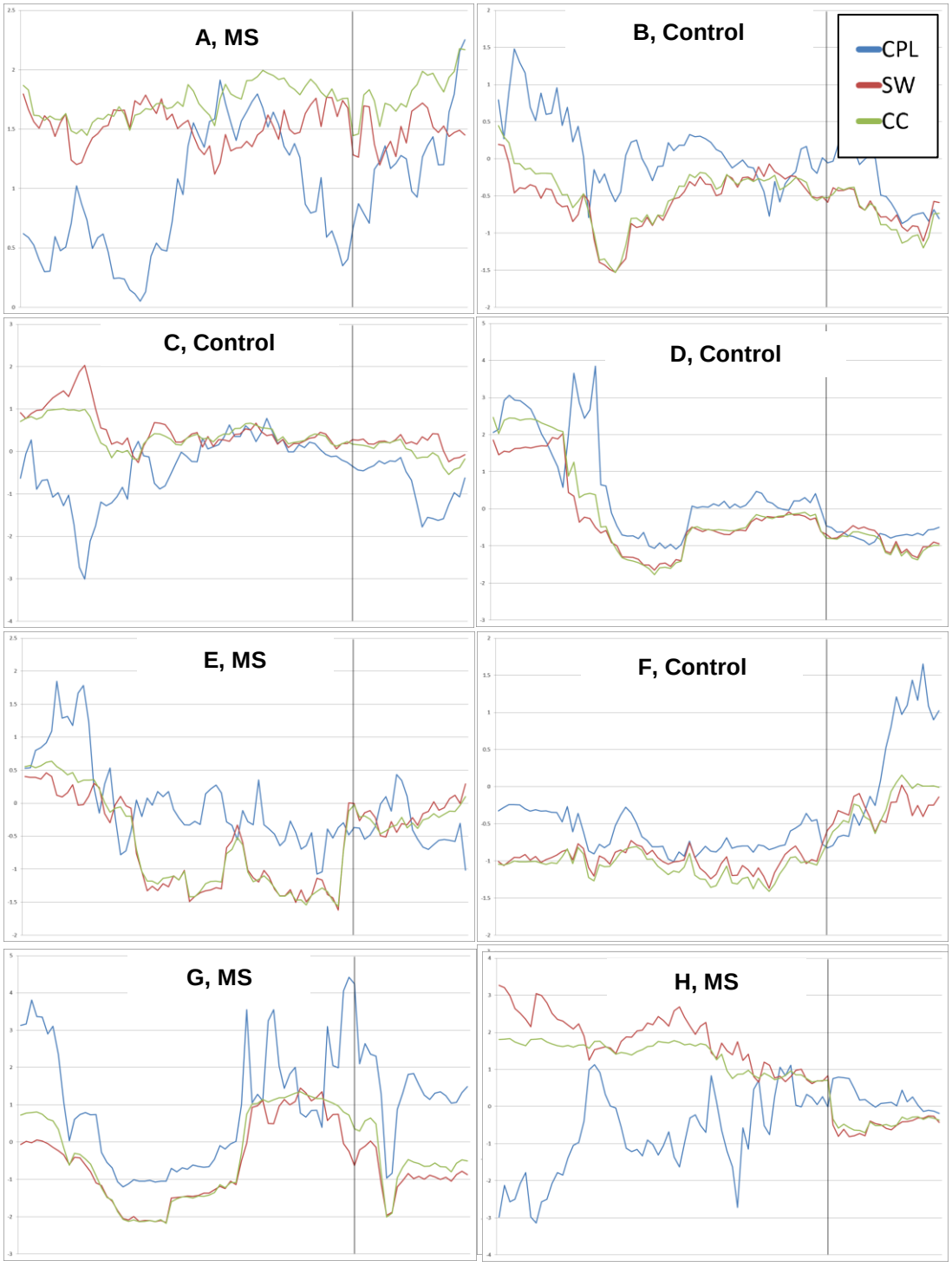
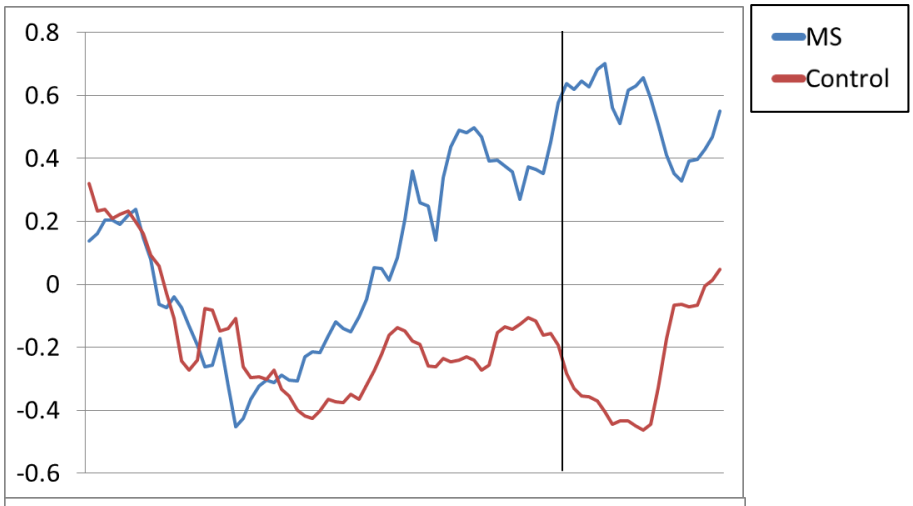
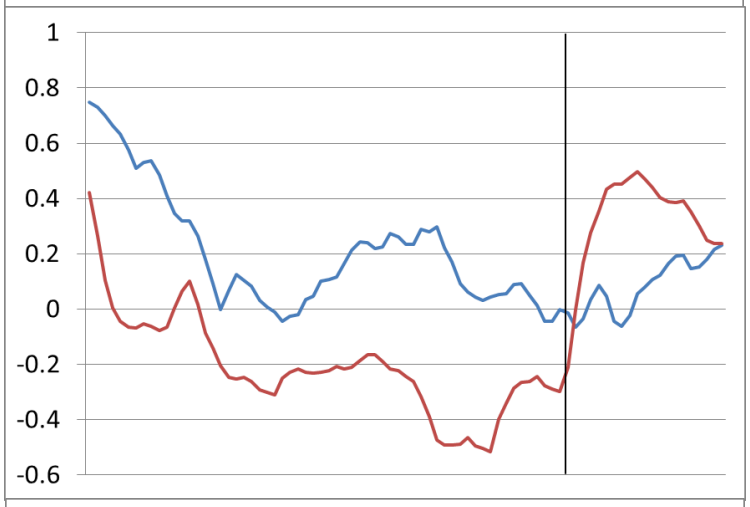


Figure 46. Metric Z-scores over time, examples from the first 8 subjects. The dark line in each graph marks the transition between task and rest periods. The x-axis represents time, with the initial 6 minute task and 2 minute rest. The y-axis represents standard deviations from the mean.

CPL


0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
MS
Control

SW


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

CC

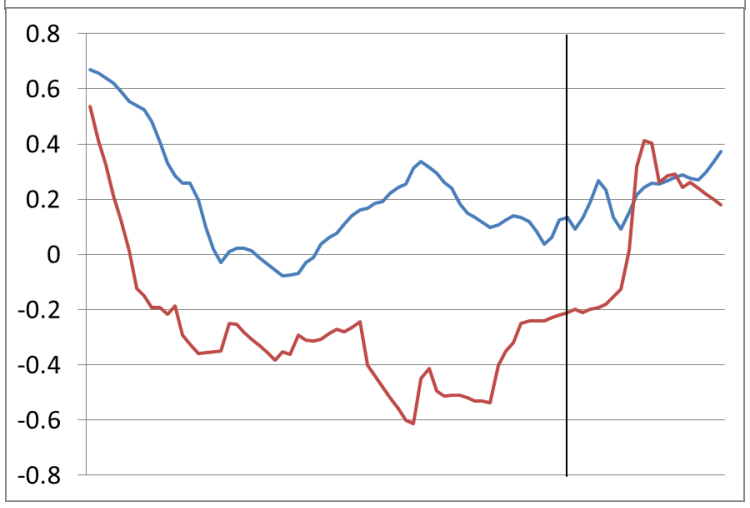

0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8

Figure 47. Mean Z-score for each metric over time. The metrics were normalised using the mean and standard deviation of all values pooled across groups, so that they would be comparable on the same graph. The dark line in each graph marks the transition between task and rest periods. The most noticeable trends common between both groups is the reduction in all metric scores at the beginning of the task period. However, there were also some differences between groups in the way the networks reorganised during and after the task. For the characteristic path length, both groups' trace fell rapidly until approximately 2 minutes into the task, at which point, where the control groups' path lengths tended to remain stable, path lengths in the MS group increased until the rest period began. For the small-worldness and clustering graphs, both groups followed a similar pattern until the rest period, where the control group saw a rapid increase in clustering and small-worldness while the MS group did not.

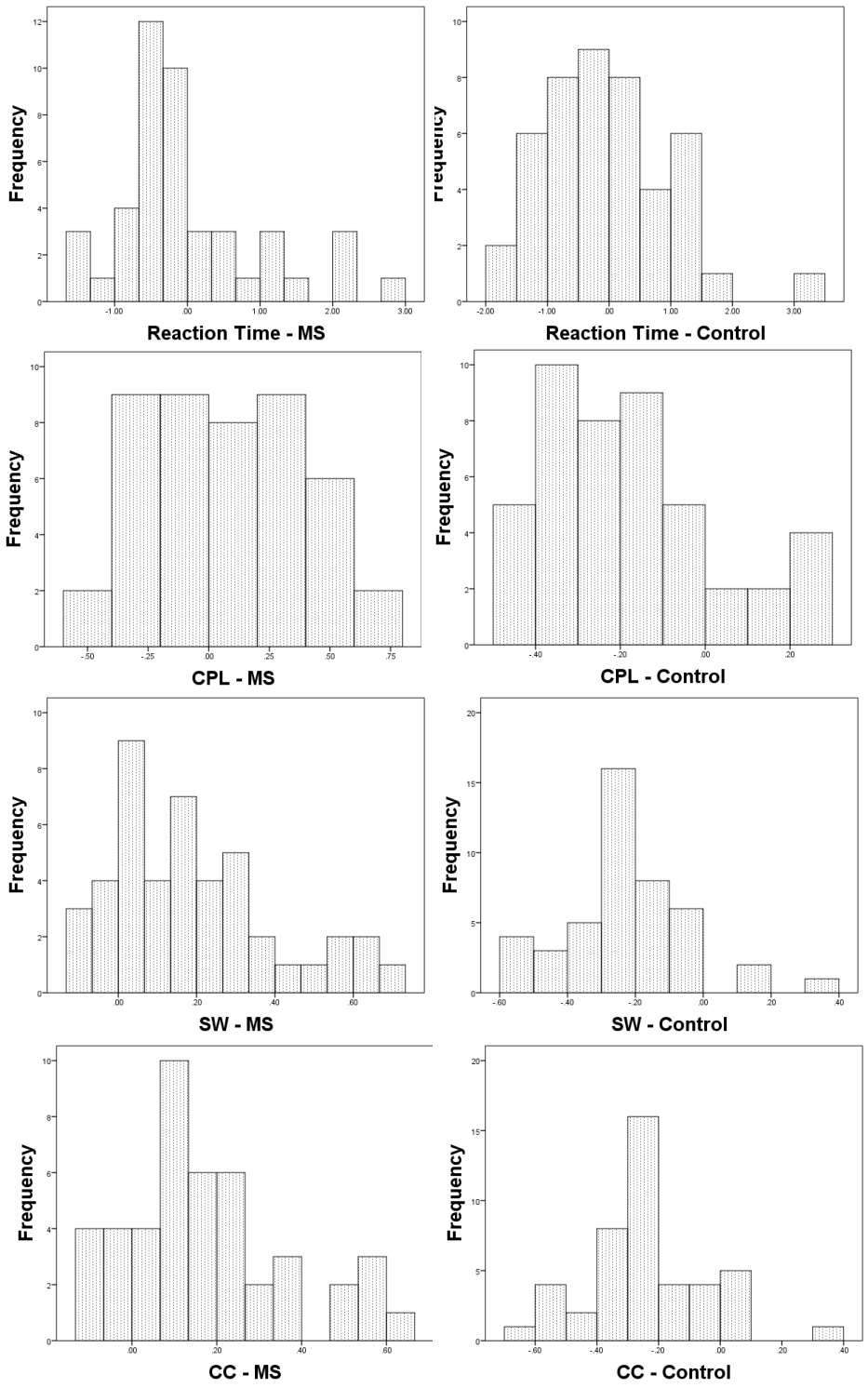


Figure 48. Histograms used to test for normality. The reaction time histograms use the mean reaction time for each subject.

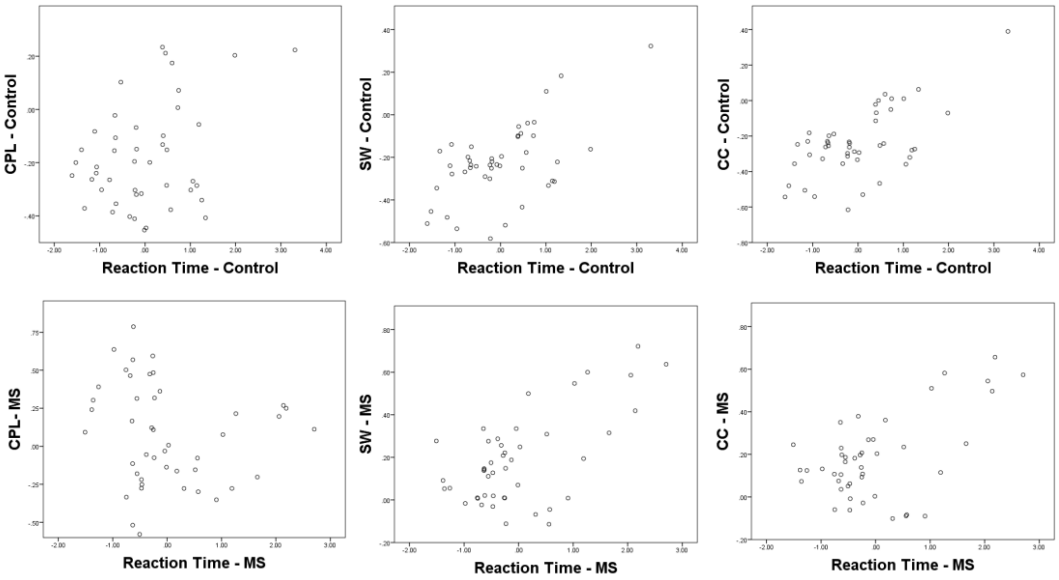


Figure 49. Correlations between mean reaction times and mean metric scores. Each datapoint represents a trial. Controls are on the top row and MS on the bottom row. For the control data set, the single outlier with a high reaction time corresponds to the first data point and, when removed, did not make the correlations with SW and CC nonsignificant.

All variables followed approximately normal distributions, as confirmed by inspection of histograms and Shapiro-Wilk tests (figure 48). Six Pearson correlation tests were performed. The 2-tailed p-value from the correlations was compared to a Bonferroni-corrected significance level of 0.008 (0.05/6). The degrees of freedom for these tests was 43 (number of pairs - 2). As the data were pooled across time and subjects, there is a likelihood of false positives.

Table 29. Outcomes of the correlation tests. Tests marked with an asterisk surpassed the Bonferroni-corrected significance threshold.

	Reaction Time / CPL (r, p)	Reaction Time / SW (r, p)	Reaction Time / CC (r, p)
Control	0.343, 0.021	0.586, <0.001 *	0.627, <0.001 *
MS	-0.154, 0.313	0.623, <0.001 *	0.554, <0.001 *

In the tests for relationships between ANT subscores and changes in network structure, there were several significant relationships. These results do not show whether there was a significant change in network structure between parts of the timecourse, but whether any change was related to the ANT subscores.

Table 30. Correlations between ANT subscores and changes in mean network metric between the final 2 minutes of the task and the 2 minutes of the subsequent rest. All values are in the format, (r, p). Asterisk indicates p-values less than 0.05.

	Characteristic Path Length	Small-Worldness	Clustering Coefficient
MS			
Alerting	-.683, 0.091	-.790, 0.034 *	-0.805, 0.029 *
Orienting	-0.569, 0.182	-0.848, 0.016 *	-0.827, 0.022 *
Conflicting	-0.961, 0.001 *	-0.699, 0.081	-0.752, 0.051
Control			
Alerting	-0.777, 0.023 *	0.228, 0.587	0.051, 0.905
Orienting	0.354, 0.390	-0.765, 0.027 *	-0.697, 0.055
Conflicting	-0.571, 0.140	0.620, 0.101	0.527, 0.180

Table 31. Correlations between ANT subscores and changes in mean network metric between the first 2 minutes of the task and the last 2 minutes of the task. All values are in the format, (r, p). Asterisk indicates p-values less than 0.05.

	Characteristic Path Length	Small-Worldness	Clustering Coefficient
MS			
Alerting	0.133, 0.776	0.059, 0.901	0.075, 0.873
Orienting	0.170, 0.715	0.517, 0.235	0.493, 0.261
Conflicting	0.132, 0.777	0.767, 0.044 *	0.723, 0.066
Control			
Alerting	-0.323, 0.436	-0.762, 0.028 *	-0.686, 0.060
Orienting	-0.452, 0.261	-0.058, 0.891	-0.240, 0.567
Conflicting	-0.665, 0.072	-0.482, 0.227	-0.587, 0.126

8.4. Discussion

In this chapter, the main aim was to investigate the responsiveness of brain network metrics to changes in network structure, which were assumed to be induced by a sustained attention task. I was able to measure three graph metrics' dynamic changes over time during the course of the task and to observe the ways in which they changed as the subjects, first, became accustomed to the task state and, second, transitioned between task and resting states. Interestingly, the patterns of the observed changes over time appeared to vary widely across subjects (figure 46), but some features were common among many subjects. For example, in most subjects there was either a dip or a peak when the rest period began in one or more graph metric. A second example is that, in most subjects, one or more metrics would dip sharply at the beginning of the task period. The plots of the mean changes

over time in metrics (figure 47) give a better idea of these features. Importantly, they show that there may be differences between MS and control groups in the way that networks behave under sustained attention. The traces for the three metrics in both groups appear to follow the same gradual decrease toward the middle of the task period (at approximately 3 minutes) and then either gradually increase or become stable. At the onset of the task, the two groups also appeared to differ, with the metric scores for the MS group not changing drastically, but the scores for the control group rising for the clustering and small-worldness metrics and dipping for the path length. These patterns could be reflecting processes by which brain networks adapt to the task demands. This may predict task performance, but for the task applied here, there were so few errors as to preclude proper analysis.

I then showed that the small-worldness and clustering metrics (which tended to follow similar patterns) were significantly related to the reaction times to trials during that block in both MS and control groups. This may be expected due to the gradual decrease in reaction time and in these metrics during task performance. It suggests that the improved reaction times may be due, in part, to short-term functional reorganisation of brain networks related to attention, visual scanning, short term memory or learning.

The extent to which the changes in network metrics over time are truly representing biological networks related to attention as opposed to confounding factors like motion due to button pressing or networks associated with the visual and motor aspects of the task is an important question. Despite correcting for motion in the image processing steps, it likely still contributes to the results, in particular during the transitions between states. However, even

if the between-state changes were due mostly to motion, this does not explain the changes that occurred within each state. While brain networks related to motor and visual functions may have contributed to the results too, it is unlikely that their contributions were the sole reason for my findings. This is evidenced by the relationships between the ANT scores and the differences in metrics between parts of the scan. The multiple significant relationships suggest that a large part of the variance in metrics between the parts of the scan is explained by performance of the attention networks in a test outside of the scanner. Another potentially important consideration is that the beginning of the final rest period coincided with the subjects closing their eyes.

A major limitation to this chapter is that, as the samples were pooled across time and subjects, they were not independent and thus the reported p-values for the tests in tables 29, 30 and 31 are highly susceptible to false positives. This is difficult to correct for, but does not invalidate observations made on the plotted data itself, although these are weaker and not supported by statistical tests. In addition, the large overlap of windows with the surorunding windows gives a smoothing effect to the outcomes.

In an improved design, the task would be designed such that it minimises the effects of motion and the impact of activation in brain networks not relevant to the task. In addition it would not involve the opening and closing of the eyes, instead using fixation. A more challenging task may be applied to increase the effect, so long as the subjects are able to understand how to perform it, as can be determined with a practice beforehand. To address a potential habituation effect, the task could be organised into regular blocks separated by rest.

Conclusions

- Although the reaction times of the MS group were slower than controls, there was no evidence of the learning effect being greater in either group.
- The graph-theoretic approach to network analysis detects dynamic changes in brain network organisation during an attention task.
- Reaction time during sustained attention in people with MS may be related to the way in which functional brain networks behave.

Part III: Discussion

9. Discussion and Conclusions

9.1. Summary of Findings

The overall aim of the thesis was to evaluate the potential usefulness of measures of brain network organisation with the graph-theoretic model of MRI data for application as markers in clinical studies of cognitive impairment and fatigue in MS. The approach taken was to assess graph-theoretic measures using 3 criteria: reliability, validity and responsiveness.

The first piece of research presented was a meta-analysis of TBSS studies of the relationship between WM tract integrity and measures of cognitive impairment and fatigue. The literature search found 12 studies with a combined 495 people with MS and 253 controls. The comparison of tract FA in people with MS and controls showed that, expectedly, FA was lower in MS for all major WM pathways, predominantly in the corpus callosum, periventricular WM and fornices. The meta-analyses for cognitive impairment and physical disability had distributions of lower FA which overlapped only minimally but, in their size and magnitude of effect, suggested a greater relevance of cerebral WM damage to cognitive rather than physical disability. These results support the argument that MS symptomatology, including cognitive symptoms, may be explained by damage to white matter and disconnection, hence supporting the applicability of connectomic analyses to the study of cognitive symptoms in MS.

Reproducibility of graph-theoretic brain network metrics was then studied (1) over time in previous studies in healthy people (2) over time in people with MS

and (3) between different sites in a travelling-subjects cohort. My data showed that reproducibility, when measured in the same sample across time, can be excellent regardless of disease status but is more difficult to achieve when additionally measured across multiple sites. Various methodological factors for reproducibility were discussed. In particular, the chosen graph density threshold range was investigated and found to be highly relevant to reproducibility.

The BraNDy-MS cohort was then described in detail. In terms of age, gender, sleep quality and depression scores, the cohort was representative of the typical RRMS and SPMS population. However, the studied sample was more impaired in cognition and fatigue measures compared to the population at large, with Z-scores being on average -0.72, -0.67 and 0.47 from the normative sample means for PASAT, SDMT and MFIS, respectively.

Significant group differences were found for all but one graph metric (ANCOVA; characteristic path length: $F=15.944$, $p=<0.001$; modularity: $F=7.063$, $p=<0.001$; small-worldness: $F=2.670$, $p=0.056$; global efficiency: $F=29.289$, $p=<0.001$; clustering coefficient: $F=3.146$, $p=0.032$). In the MS group, linear regression analyses showed that cognitive impairment was strongly predicted by a combination of graph metrics and demographic variables (predicting 55.6% of variance in the PASAT and 60.0% of the variance in the SDMT), but fatigue measures were not. None of the regression models predicted the cognitive or fatigue variables in the control group.

Reaction times to the mSDMT task were slower on average in the MS group than in controls, but did not differ in rate of change, suggesting that any learning or fatiguing effects were similar in both groups. Dynamic changes in graph metrics were observed during the sustained attention task using the graph-theoretic approach, and these changes did differ between groups, but may have been due to a habituation effect. Changes were observed at the transition between task and rest periods, but also within the task period.

9.2. Are Graph-Theoretic Brain Network Metrics Good Candidate Markers of Cognition or Fatigue in MS?

The research presented in this thesis gives evidence both for and against graph metrics being considered good markers of cognition or fatigue. First, for graph metrics based on functional data, although the regression models for the two fatigue variables were significant (MFIS: $F=6.259$, $p=<0.001$, $R^2=0.629$; NFI-MS: $F=2.989$, $p=0.015$, $R^2=0.576$), the prediction was not better than that in the demographic variables-only block (MFIS: $F=11.716$, $p=<0.001$, $R^2=0.693$; NFI-MS: $F=4.349$, $p=0.005$, $R^2=0.446$). In the combined models with graph metrics entered, the variables which contributed significantly to the prediction for fatigue did not include any network metrics, except for the MFIS regression having the modularity score surpass the threshold. In comparison to the cognitive measures which had several graph metric predictors each being significant, this amounts to only a small contribution. The dynamic network analyses may provide some limited evidence of the contrary in the fact that the brain networks did alter differently during the task performance compared to controls. Assuming that this is a

fatiguing effect (which, however, was not observed in the reaction time scores), it may suggest a relationship, but this is speculative. Together, this evidence suggests that fatigue is better predicted by the standard questionnaires than by brain network organisation. No previous studies had focussed on investigating the relationship between brain network organisation and fatigue, instead focussing on cognition.

For cognition, my findings were more positive. First, the meta-analysis showed strong evidence for the relationship between tract integrity and cognition which, while not directly relevant to graph metrics, does support the idea that the graph theoretic method could be a useful and appropriate way to model network disruption in the MS brain. With the sample leaning toward being more cognitively-impaired than the population, the results are likely more generalisable to that group. The two main tests of cognition, the PASAT and the SDMT, were strongly predicted by a combination of graph metrics and demographic variables (PASAT: $F=6.452$, $p=<0.001$, $R^2=0.773$; SDMT: $F=4.721$, $p=0.002$, $R^2=0.724$). The graph metrics which contributed to the prediction all pointed to less efficient information transfer. Of the ANT subscores, only the score for alerting reached the significance threshold. This may suggest a greater relevance of network organisation for more complex tasks requiring fast information processing than for simple attention or reaction-time tasks.

In terms of reproducibility, there is good evidence across multiple independent studies that reproducibility of graph metrics is adequate for clinical studies, but that care should be taken when combining data from multiple sources. I have further shown that disease status is not likely a large obstacle to

reproducibility. My findings suggest that, while good reproducibility is possible, it requires an understanding of the graph theoretic method and its parameters as well as the correct processing steps to be taken.

My investigation of the responsiveness criterion was more exploratory and the first of its kind, to my knowledge. While I did show that graph metrics do change in response to the transition between states and in the performance of the task, the direction and nature of these changes were variable, and might be highly dependent on individual differences in functional network structure or on the cognitive strategy applied to the task. The significant correlations between graph metrics over time and reaction times do indicate that the dynamic changes in network organisation were relevant to the task. However, there were multiple drawbacks to this study, mostly in distinguishing real biological network change from motion or motor and visual activation. Evidence for the importance of responsiveness to cognition or fatigue from this analysis was relatively weak, but did serve to provide some possible good avenues of future investigation.

Criteria for the practical implementation of graph metrics as markers of cognition include ease of interpretation, time efficiency and cost efficiency. First, although an MRI scan is more expensive and time-consuming than completing a questionnaire or neuropsychological test, MRI is often already performed to gather information on lesion load, atrophy and other conventional markers of brain health in MS. The addition of an fMRI scan (5-10 minutes) is a relatively low cost when the MRI is already being performed.

With this evidence in balance, I can recommend graph-theoretic brain network metrics for use as markers of cognition in future research of MS. For clinical trials, the network metrics need to be shown to relate to cognitive scores on an individual level, rather than on a group level.

9.3. Limitations

All models make assumptions. Assumptions are simplifications which allow us to ignore certain factors about a real-world object which we deem negligible or unimportant for our particular aims. George Box said, “essentially, all models are wrong but some are useful” [347]: no model can or should completely capture every aspect of the thing being modelled, but on the other hand, while models are approximations, they don’t have to be “correct” to be useful. In the context of brain networks, there are a number of assumptions made in graph analysis which are therefore important to acknowledge. First, graphs assume that their nodes, while not necessarily separate, are mutually distinct. In other words, that each one encapsulates exclusively some part of the modelled network. For example, while the motor and supplementary motor areas of the brain are considered functionally-distinct, they are anatomically adjacent and directly connected by WM fibres. To translate that assumption to the case of the brain is a simplification because the brain is not neatly divided into segments, in either a functional or an anatomical sense. Second, while the whole graph is considered separate from its environment, external factors do play a large role in the function of the brain. It is not clear where the boundary of the system lies, but is assumed to be at the edge of the cerebrum. Another relevant example which illustrates the point is that network efficiency

measures depend on the assumption that the shortest path is always followed from node A to node B, but this is not always true of the brain. The graph-network distinction is therefore important to bear in mind in studies of this type.

Other limitations are found in the characteristics of the sample. First, while the decision to include individuals with MS who were using DMTs was based on reducing selection bias, the effects of DMTs on cognitive function are not yet understood. The 37% of the MS group on DMTs may have therefore altered the validity of the results. On one hand, the sample may be more representative of the population, as a similar proportion are also likely taking DMTs, but, on the other hand, the individuals on DMTs may be affected by the therapeutic action of the drug. Second, since the volunteers recruited through the CRAMMS trial were screened for cognitive impairment, they tended to be more cognitively impaired than the volunteers recruited through the MS clinics or elsewhere. Twenty-five of the 37 people with MS (68%) were recruited via CRAMMS and 12 were not. This skewed the sample toward being more cognitively impaired than the normative samples used in comparison (in the PASAT by 0.72 standard deviations below the mean of the normative sample and in the SDMT by 0.62 standard deviations below the mean of the normative sample, both indicating worse performance). Having said that, the normative samples were drawn from the North American population and may therefore differ from the BraNDy cohort in multiple demographic and socioeconomic factors. Again, this alters the generalisability of my results toward the cognitively impaired population moreso than the general MS population.

A drawback to using graph-theory to produce summary measures in people with MS is that the model only applies to the brain as a whole. Because of the range of symptoms and variation in distributions of lesions in MS, while significant differences exist between groups, it may not be a useful approach for understanding individual cases. In addition, in cases where a single structure with a high importance for cognition is damaged by a focal lesion, the graph-theoretic model may not be suitable as it could mask these types of effects due to its whole-brain perspective.

9.4. Future Directions

One issue with studies applying the graph-theoretic method is in the interpretation of graph metrics in the context of the brain. There is potential for future research into what entails a good or bad change in network organisation, and how good or bad should even be defined in this context. For example, is an increased nodal centrality score a good or a bad thing?

In studies of brain network topology in neurologic disease, findings of both increased and decreased network efficiency have been reported in relation to cognitive impairment. On one hand, authors explain reduced efficiency in MS as being a result of damage to brain networks which disrupts key network hubs and therefore the overall flow of information [184]. On the other, where networks appear to have become more efficient in the disease group, the explanation given is that the brain has undergone plastic change sufficient to compensate for any neurologic damage [186]. These explanations are not mutually exclusive and both processes could be happening simultaneously. Further research should attempt to disentangle these processes. In MS

research, a discussion about adaptive and maladaptive changes is taking place, but needs to be supported by neuroimaging evidence.

One relatively unexplored avenue of investigation in MS research is that of network robustness and cognitive reserve [176], where it could be speculated that a graph which is more robust to damage corresponds to a healthier brain network and has a higher threshold for clinically-relevant impairment [170]. However, testing this may require longitudinal data from people with radiologically-isolated syndromes.

While the results of the responsiveness analysis show that the graph metrics are responsive to changes in cognitive state, it remains to be seen whether the method is sensitive enough to detect changes in network organisation in the long-term in response to cognitive rehabilitation. It is important to demonstrate this in longitudinal data for studies which aim to monitor response to an intervention.

9.5. Conclusions

- Cerebral WM damage is relevant to cognitive impairment in MS.
- Graph theory is an appropriate way to model brain network disruption in MS.
- Graph metrics are reproducible across time and in people with MS, but less so across multiple sites.
- Brain network organisation is significantly different in people with MS from matched healthy controls.

- Graph metrics are significantly associated with measures of cognition in people with MS, in particular those involving complex information processing.
- The brain networks of people with MS are more segregated and less integrated, implying less efficient information transfer and lower network small worldness.
- Graph metrics are responsive to induced changes in cognitive state in the immediate term.
- Graph-theoretic brain network metrics are suitable candidate markers of cognition in MS, but not for fatigue.

Appendices

A. Retrieving local maxima from TBSS t-statistic maps

```
# Retrieve the local positive peaks
cluster -i ${1} --mm -n 100000 --olmax=${1}_pos.txt -t 0.1

# Create a negative image to subsequently extract negative
peaks
fslmaths ${1} -mul -1 ${1}_neg.nii.gz

# Retrieve the local negative peaks
cluster -i ${1}_neg --mm -n 100000 --olmax=${1}_neg.txt -t 0.1

# Remove the negative image
rm ${1}_neg.nii.gz
```

B. Research Ethics Committee Approval Letter

NHS
Health Research Authority
NRES Committee East Midlands - Nottingham 2
The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

Telephone: 0115 8839425

24 March 2014

Dr Rob Dineen
Clinical Associate Professor of Neuroimaging
University of Nottingham
Room B1435 Queen's Medical Centre

Dear Dr Dineen,

Study title:	**Imaging Markers of Brain Network Dysfunction in Multiple Sclerosis**
REC reference:	**14/EM/0064**
Protocol number:	**14010**
IRAS project ID:	**141427**

Thank you for your letter of 07 March 2014, responding to the Committee's request for further information on the above research and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Vice-Chair.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the REC Manager - Liza Selway, on 0115 8839425.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised, subject to the conditions specified below.

Ethical review of research sites

NHS sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Non-NHS sites

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database within 6 weeks of recruitment of the first participant (for medical device studies, within the timeline determined by the current registration and publication trees).

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non clinical trials this is not currently mandatory.

If a sponsor wishes to contest the need for registration they should contact Catherine Blewett (catherineblewett@nhs.net), the HRA does not, however, expect exceptions to be made. Guidance on where to register is provided within IRAS.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	*Version*	*Date*
Advertisement	POSTER	23 January 2014
Covering Letter		28 January 2014
Covering Letter		07 March 2014
Evidence of insurance or indemnity		31 July 2013
Letter from Sponsor		23 January 2014
Letter of invitation to participant	1	23 January 2014
Other: NON--NHS site - SSI form		28 January 2014
Other: Summary CV - supervisor 2	DA	28 January 2014
Other: Summary CV - supervisor 3	CC	
Other: Cognitive ANT Test	1	28 January 2014
Other: Cognitive BDUI Test	1	28 January 2014
Other: Cognitive MFIS	1	28 January 2014
Other: Cognitive MSFC Test	1	28 January 2014
Other: Cognitive PSQI Test		28 January 2014
Other: Cognitive SDMT Test		28 January 2014
Other: CV Chief Investigator		
Other: CV Student		
Other: Letter from Funder		05 August 2013
Participant Consent Form: MS	1.1	03 March 2014
Participant Consent Form: Healthy Volunteer	1.1	03 March 2014
Participant Information Sheet: MS	1.1	03 March 2014
Participant Information Sheet: Healthy Volunteer	1.2	07 March 2014
Protocol	1	23 January 2014
REC application	141427/5566 99/1/16	29 January 2014
Referees or other scientific critique report	Response to Feedback	
Referees or other scientific critique report	Lay & Peer review feedback	
Response to Request for Further Information		07 March 2014

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document *"After ethical review – guidance for researchers"* gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

14/EM/0064	Please quote this number on all correspondence

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

Yours sincerely

Dr Simon Roe
Vice-Chair

Email: NRESCommittee.EastMidlands-Nottingham2@nhs.net

Enclosures: "After ethical review – guidance for researchers" *SL-AR2*

Copy to: *Paul Cartledge, University of Nottingham*
Charlotte Davies, Nottingham University Hospitals NHS Trust

C. Consent Form

The University of Nottingham
UNITED KINGDOM • CHINA • MALAYSIA

CONSENT FORM
(Final version 1.0: 23/01/14)

Title of Study: Imaging Markers of Brain Network Dysfunction in Multiple Sclerosis

REC ref: *14/EM/0064*

Name of Researcher: Dr RA Dineen, Prof DP Auer, Prof C Constantinescu, T Welton, A French

Name of Participant: **Please initial box**

1. I confirm that I have read and understand the information sheet version number ……dated..... for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my medical care or legal rights being affected. I understand that should I withdraw then the information collected so far cannot be erased and that this information may still be used in the project analysis. ☐

3. I understand that relevant sections of my medical notes and data collected in the study may be looked at by authorised individuals from the University of Nottingham, the research group and regulatory authorities where it is relevant to my taking part in this study. I give permission for these individuals to have access to these records and to collect, store, analyse and publish information obtained from my participation in this study. I understand that my personal details will be kept confidential. ☐

4. I agree to comply with the reasonable instructions of the supervising investigator and will notify him immediately of any unexpected unusual symptoms or deterioration of health. ☐

5. I understand and agree to undergo one or two MRI scans (as required), separated by a period of approximately 28 days and a cognitive assessment for analysis of brain network dysfunction, cognitive dysfunction and fatigue. ☐

6. In the event of the imaging showing a significant incidental abnormality, I authorise the investigators to disclose the details of the findings to me and to my general practitioner, and will provide details of the scan findings to them so that any further investigations or referrals can be made. ☐

7. **(Optional)** I agree for data collected about me to be used in an anonymised form in future pooled analyses within the University of Nottingham. ☐

6. I agree to take part in the above study. ☐

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Name of Person taking consent	Date	Signature

3 copies: 1 for participant, 1 for the project notes and 1 for the medical notes

D. Information Sheet

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Faculty of Medicine & Health Sciences
School of Medicine
Radiological Sciences Research Group
The University of Nottinham
University Park
Nottingham
NG7 2RD

t: +44 (0)115 823 0018
f: +44 (0)115 823 0004
e: medicine@nottingham.ac.uk

www.nottingham.ac.uk/medicine

Imaging Markers of Brain Network Dysfunction in Multiple Sclerosis

Participant Information Sheet

Name of Researcher(s): Dr RA Dineen, Prof DP Auer, Prof C Constantinescu, T Welton, A French

We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. One of our team will go through the information sheet with you and answer any questions you have. Talk to others about the study if you wish. Ask us if there is anything that is not clear.

What is the purpose of the study?

In medical science, a 'marker' is something that can be measured (for example in the blood or urine, or on a scan) to monitor a biological process, disease or treatment response. In its 2013-17 Research Strategy, the multiple sclerosis (MS) Society calls for the development of better imaging markers that can be used in trials of new treatments for MS and MS symptoms. As new treatments are developed for treating problems with thought processes such as memory and concentration (referred to as cognitive processes) in MS, it is important to have a meaningful and easily-interpretable marker to show how organisation of brain networks alters in response to the new treatments.

We propose that a type of analysis called 'Graph Theory' (GT) will be able to provide meaningful and easily interpretable imaging markers of brain organisation from functional magnetic resonance images (MRI). GT has been applied widely to study complex networks in science and society - for example, to characterise worldwide social networks on the internet, such as Facebook. The purpose of the study is therefore to investigate whether or not markers derived from GT analysis relate to cognitive processes, and how stable those markers are over time.

Why have I been invited?

As you have been diagnosed with multiple sclerosis, you are being invited to take part. You were selected from a list of individuals who had agreed to take part in research. We selected you based on your age, MS subtype, and cognitive ability. We are inviting 60 participants like you to take part: 30 people with MS and 30 people without MS.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. This would not affect your legal rights.

What will happen to me if I take part?

Participation requires at most two attendances, the first lasting approximately 2 hours and the second (if required) approximately 1 hour. A researcher will contact you to arrange times that are convenient. Three days before each session, a researcher will contact you by telephone to confirm the time, and to check that you are well-enough to take part.



The MRI scanner at QMC that will be used for the study

For the first attendance, a researcher will meet you at the entrance to the Queen's Medical Centre. You will go to a private room where the researcher will explain the study to you and answer any questions you have. If you have decided you would like to take part, you will be asked to sign a consent form. The researcher will go through a set of questionnaires with you to test various cognitive processes. You will then go to the MRI scanner inside the Queen's Medical Centre, where a researcher will check that it is safe for you to be scanned. You will then have an MRI scan of your brain, lasting approximately 40 minutes. Inside the scanner, you will simply have to lie down and relax without falling asleep.

The second attendance will be roughly 28 days after the first. It will be the same as the first session, but without the questionnaires.

Expenses and payments

Participants will be reimbursed £10 for travel, parking and inconvenience caused as a result of taking part in the study.

What are the possible disadvantages and risks of taking part?

Provided you do not have a condition which prevents you from having an MRI scan, there are no risks associated with this study. You will be screened for conditions preventing you from having an MRI scan before you have the scan.

Magnetic resonance imaging uses radio waves similar to those used in radio and TV transmission. These have a much lower energy than X-rays and are therefore considered safe to humans. We will be following strict national safety guidelines which are designed to prevent the theoretical hazards of MRI which are burns and electric shocks. Such accidents have very rarely occurred elsewhere. While there is no evidence to suggest that MRI is harmful during pregnancy, it is advised not to scan pregnant women. We have decided not to test for pregnancy as routine but if you think you may be pregnant you should not be scanned.

The MRI scanner is a relatively enclosed space and occasionally participants can feel claustrophobic. During the scan you are able to speak to the researchers performing the scan. If you would like to come out of the scanner at any time, you can request this. If you know that you are claustrophobic, we would advise against you participating.

In the unlikely event that a potential abnormality is identified on the scan by one of the researchers, the images will be reviewed by a qualified consultant radiologist working within the Radiological Sciences Research Group. If the suspicion of a significant abnormality is confirmed, the findings will be explained to you by the radiologist and details of the findings sent to your hospital neurologist. The radiologist involved will advise on appropriate measures which will need to be taken in light of the findings, which may include performing further scans or referral to an appropriate hospital-based specialist, both of which would be arranged by your hospital neurologist. The finding of a significant unexpected abnormality may have benefits in that it may be possible to offer treatments earlier than would otherwise have been possible. However, participants should also be aware that the finding of significant abnormality may adversely affect employment and insurance status.

What are the possible benefits of taking part?

We cannot promise the study will help you personally, but if it provides a marker that can be successfully employed in trials targeting cognitive symptoms in MS, it is possible that in future, improvements will be made to the management of such symptoms, including fatigue.

In terms of producing knowledge, we anticipate the resulting publication of two scientific papers, with one of potentially high-impact. The work completed in the study as part of an educational programme will increase the capacity for future MS research, by helping to train capable researchers, and by developing and improving partnerships between the involved parties (such as those between research institutions, the health sector and local MS communities). The study results could also inform future product development.

What happens when the research study stops?

Upon completing the research study, you will not be contacted further (unless you wish to hear the results of the study). Data will be analysed and kept on a secure server for 7 years after the end of the study. Scientific manuscripts will be prepared and submitted to journals for dissemination.

What if there is a problem?

In case you have a complaint on your treatment by a member of staff or anything to do with the study, you can initially approach the lead investigator. If this route fails to achieve a satisfactory resolution and you still wish to complain about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms may be available to you. The Patient Advice and Liaison Service (PALS) can be contacted for further assistance at the QMC by calling 0800 1830204.

Will my taking part in the study be kept confidential?

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. Any information about you which leaves the research unit will have your name and address removed so that you cannot be recognised from it. Your personal details will not be passed onto any third parties.

If you join the study, some parts of your medical records and the data collected for the study will be looked at by authorised persons from the University of Nottingham who are organising the research. They may also be looked at by authorised people to check that the study is being carried out correctly. All will have a duty of confidentiality to you as a research participant and we will do our best to meet this duty.

All information which is collected about you during the course of the research will be kept strictly confidential, stored in a secure and locked office, and on a password protected database. Any information about you which leaves the hospital will have your name and address removed (anonymised) and a unique code will be used so that you cannot be recognised from it.

Your personal data (address, telephone number, email) will be kept for 1 year after the end of the study so that we are able to contact you about the findings of the study (if you inform us that you do not object to being contacted). After this time your data will be disposed of securely. All other data (research data) will be kept securely for 7 years. During this time it may be used in an anonymised form in future brain imaging studies but all precautions will be taken by all those involved to maintain your confidentiality, only members of the research team will have access to your personal data. Should you disclose anything to us which we feel puts you or anyone else at any risk, we may feel it necessary to report this to the appropriate persons.

What will happen if I don't want to carry on with the study?

Your participation is voluntary and you are free to withdraw at any time, without giving any reason, and without your legal rights being affected. If you withdraw then the information collected so far cannot be erased and this information may still be used in the project analysis. No further data would be collected or any other research procedures carried out on or in relation to you.

What will happen to the results of the research study?

The MRI scans and questionnaire scores will be used by researchers in the Radiological Sciences Research Group or collaborating researchers to understand how brain network organisation relates to cognitive performance and fatigue in MS. The results will be written-up as part of a PhD thesis and published in peer-reviewed scientific journals. The results of the study are likely to be published between 2014 and 2016 in journals such as *Neurology*, *Brain* and *Human Brain Mapping*. In addition, the MS Society regularly disseminates news about the results of the research they fund. You will not be personally identified in any publication.

Who is organising and funding the research?

This research is being organised by the Radiological Sciences Research Group in the University of Nottingham and is being funded by the UK Multiple Sclerosis Society.

Who has reviewed the study?

All research in the NHS is looked at by an independent group of people, called a Research Ethics Committee, to protect your interests. This study has been reviewed and given favourable opinion by the "East Midlands – Nottingham 2" NHS Research Ethics Committee. We performed a pilot study in 10 participants with MS, who all agreed that the scanning and assessment were acceptable. The chief investigator attended a charity-run MS support

centre and was able to discuss the research idea with the centre manager and present individuals with MS.

Further information and contact details

If you would like to discuss the study further or would like more information, please feel free to contact the chief investigator:

Dr Rob Dineen,
Clinical Associate Professor,
Radiological Sciences Research Group,
Division of Clinical Neuroscience,
Queen's Medical Centre,
Nottingham,
NG7 2UH

+44 (0)115 8231173
rob.dineen@nottingham.ac.uk

General information and useful links regarding participation in clinical research is provided by the People in Research website: http://www.peopleinresearch.org

The Patient Advice and Liaison Service (PALS) can be contacted for further independent assistance, or regarding any complaints at the QMC by calling 0800 1830204.

References

1. Mackenzie, I.S., et al., *Incidence and prevalence of multiple sclerosis in the UK 1990-2010: a descriptive study in the General Practice Research Database.* J Neurol Neurosurg Psychiatry, 2014. **85**(1): p. 76-84.
2. Rosati, G., *The prevalence of multiple sclerosis in the world: an update.* Neurol Sci, 2001. **22**(2): p. 117-39.
3. Rao, S.M., et al., *Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction.* Neurology, 1991. **41**(5): p. 685-91.
4. Schwid, S.R., et al., *Fatigue in multiple sclerosis: current understanding and future directions.* J Rehabil Res Dev, 2002. **39**(2): p. 211-24.
5. Kornblith, A.B., N.G. La Rocca, and H.M. Baum, *Employment in individuals with multiple sclerosis.* Int J Rehabil Res, 1986. **9**(2): p. 155-65.
6. Gregory, R.J., P. Disler, and S. Firth, *Employment and Multiple Sclerosis in New Zealand.* Journal of Occupational Rehabilitation, 1993. **3**(2): p. 113-117.
7. Fineberg, N.A., et al., *The size, burden and cost of disorders of the brain in the UK.* J Psychopharmacol, 2013. **27**(9): p. 761-70.
8. Sobocki, P., et al., *Estimation of the cost of MS in Europe: extrapolations from a multinational cost study.* Mult Scler, 2007. **13**(8): p. 1054-64.
9. Trisolini, M., et al., *Global economic impact of multiple sclerosis.* London (United Kingdom): Multiple Sclerosis International Federation, 2010.
10. Kister, I., et al., *Natural History of Multiple Sclerosis Symptoms.* International Journal of MS Care, 2013. **15**(3): p. 146-158.
11. Peyser, J.M., et al., *Guidelines for neuropsychological research in multiple sclerosis.* Arch Neurol, 1990. **47**(1): p. 94-7.
12. Benedict, R.H., et al., *Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS).* J Int Neuropsychol Soc, 2006. **12**(4): p. 549-58.
13. Lerdal, A., E.G. Celius, and T. Moum, *Fatigue and its association with sociodemographic variables among multiple sclerosis patients.* Mult Scler, 2003. **9**(5): p. 509-14.
14. Rupprecht, S., et al., *Prevalence of Sleep Disturbances, Fatigue, Anxiety and Depression in Multiple Sclerosis (MS) (P1.128).* Neurology, 2015. **84**(14 Supplement).
15. Weiland, T.J., et al., *Clinically Significant Fatigue: Prevalence and Associated Factors in an International Sample of Adults with Multiple Sclerosis Recruited via the Internet.* PLOS ONE, 2015. **10**(2): p. e0115541.
16. Chiaravalloti, N.D. and J. DeLuca, *Cognitive impairment in multiple sclerosis.* The Lancet Neurology, 2008. **7**(12): p. 1139-1151.

17. Rahn, K., B. Slusher, and A. Kaplin, *Cognitive Impairment in Multiple Sclerosis: A Forgotten Disability Remembered.* Cerebrum: the Dana Forum on Brain Science, 2012. **2012**: p. 14.
18. He, D., et al., *Pharmacological treatment for memory disorder in multiple sclerosis.* Cochrane Database Syst Rev, 2013(12): p. CD008876.
19. Messinis, L., et al., *Assessment and rehabilitation of cognitive impairment in multiple sclerosis.* Int Rev Psychiatry, 2010. **22**(1): p. 22-34.
20. Taus, C., et al., *Amantadine for fatigue in multiple sclerosis.* Cochrane Database Syst Rev, 2003(2): p. CD002818.
21. Foley, F.W., et al., *The Need for Screening, Assessment, and Treatment for Cognitive Dysfunction in Multiple Sclerosis: Results of a Multidisciplinary CMSC Consensus Conference, September 24, 2010.* International Journal of MS Care, 2012. **14**(2): p. 58-64.
22. Benedict, R.H.B. and K. Fuchs, *Introduction: Cognitive Dysfunction in Multiple Sclerosis.* International Journal of MS Care, 2012. **14**(2): p. 55-57.
23. Larson, R.D., *Psychometric Properties of the Modified Fatigue Impact Scale.* International Journal of MS Care, 2013. **15**(1): p. 15-20.
24. Mills, R.J., et al., *Development of a patient reported outcome scale for fatigue in multiple sclerosis: The Neurological Fatigue Index (NFI-MS).* Health Qual Life Outcomes, 2010. **8**: p. 22.
25. Smith, A., *Symbol digit modalities test: Manual*. 2013.
26. Tombaugh, T.N., *A comprehensive review of the Paced Auditory Serial Addition Test (PASAT).* Arch Clin Neuropsychol., 2006. **21**(1): p. 53-76.
27. Barkhof, F., *The clinico-radiological paradox in multiple sclerosis revisited.* Curr Opin Neurol, 2002. **15**(3): p. 239-45.
28. Enzinger, C. and F. Fazekas, *Measuring Gray Matter and White Matter Damage in MS: Why This is Not Enough.* Frontiers in Neurology, 2015. **6**: p. 56.
29. Goodin, D.S., *Magnetic resonance imaging as a surrogate outcome measure of disability in multiple sclerosis: Have we been overly harsh in our assessment?* Annals of Neurology, 2006. **59**(4): p. 597-605.
30. Fox, R.J., et al., *Setting a research agenda for progressive multiple sclerosis: the International Collaborative on Progressive MS.* Mult Scler, 2012. **18**(11): p. 1534-40.
31. MS Society Research Strategy Committee, *MS Society Research Strategy 2013-17*. 2013, London.
32. Bullmore, E. and O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems.* Nat Rev Neurosci, 2009. **10**(3): p. 186-198.
33. Filippi, M., et al., *Assessment of system dysfunction in the brain through MRI-based connectomics.* The Lancet Neurology, 2013. **12**(12): p. 1189-1199.
34. Schoonheim, M.M., K.A. Meijer, and J.J. Geurts, *Network collapse and cognitive impairment in multiple sclerosis.* Front Neurol, 2015. **6**: p. 82.
35. Dineen, R.A., et al., *Disconnection as a mechanism for cognitive dysfunction in multiple sclerosis.* Brain, 2009. **132**(Pt 1): p. 239-49.

36. Welton, T., et al., *Functionally Relevant White Matter Degradation in Multiple Sclerosis: A Tract-based Spatial Meta-Analysis.* Radiology, 2015. **275**(1): p. 89-96.
37. Bo, L., et al., *Grey matter pathology in multiple sclerosis.* Acta Neurol Scand Suppl, 2006. **183**: p. 48-50.
38. Fleischer, V., et al., *Increased structural white and grey matter network connectivity compensates for functional decline in early multiple sclerosis.* Multiple Sclerosis Journal, 2017. **23**(3): p. 432-441.
39. Fitzpatrick, R., et al., *Evaluating patient-based outcome measures for use in clinical trials.* 1998.
40. Mayeux, R., *Biomarkers: Potential Uses and Limitations.* NeuroRx, 2004. **1**(2): p. 182-188.
41. Coster, W.J., *Making the Best Match: Selecting Outcome Measures for Clinical Trials and Outcome Studies.* The American Journal of Occupational Therapy, 2013. **67**(2): p. 162-170.
42. Jerosch-Herold, C., *An Evidence-Based Approach to Choosing Outcome Measures: A Checklist for the Critical Appraisal of Validity, Reliability and Responsiveness Studies.* British Journal of Occupational Therapy, 2005. **68**(8): p. 347-353.
43. Clanet, M., *Jean-Martin Charcot. 1825 to 1893.* Int MS J, 2008. **15**(2): p. 59-61.
44. Compston, A., et al., *McAlpine's Multiple Sclerosis*. 4th ed. 2005, Philadelphia: Churchill Livingstone.
45. Handunnetthi, L., et al., *Regulation of major histocompatibility complex class II gene expression, genetic variation and disease.* Genes Immun, 2010. **11**(2): p. 99-112.
46. Alcina, A., et al., *Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations.* PLoS One, 2012. **7**(1): p. e29819.
47. Zhang, R., et al., *Association between the IL7R T244I polymorphism and multiple sclerosis: a meta-analysis.* Mol Biol Rep, 2011. **38**(8): p. 5079-84.
48. Ramagopalan, S.V., et al., *Rare variants in the CYP27B1 gene are associated with multiple sclerosis.* Ann Neurol, 2011. **70**(6): p. 881-6.
49. Sadovnick, A.D., et al., *Evidence for genetic basis of multiple sclerosis. The Canadian Collaborative Study Group.* Lancet, 1996. **347**(9017): p. 1728-30.
50. Ascherio, A. and K.L. Munger, *Environmental risk factors for multiple sclerosis. Part I: the role of infection.* Ann Neurol, 2007. **61**(4): p. 288-99.
51. Orton, S.M., et al., *Sex ratio of multiple sclerosis in Canada: a longitudinal study.* Lancet Neurol, 2006. **5**(11): p. 932-6.
52. Simpson, S., Jr., et al., *Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis.* J Neurol Neurosurg Psychiatry, 2011. **82**(10): p. 1132-41.
53. Hernan, M.A., M.J. Olek, and A. Ascherio, *Geographic variation of MS incidence in two prospective studies of US women.* Neurology, 1999. **53**(8): p. 1711-8.

54. Farez, M.F., et al., *Sodium intake is associated with increased disease activity in multiple sclerosis.* J Neurol Neurosurg Psychiatry, 2014. **86**: p. 26-31.
55. Benedict, R.H. and R. Zivadinov, *Risk factors for and management of cognitive dysfunction in multiple sclerosis.* Nat Rev Neurol, 2011. **7**(6): p. 332-42.
56. Handel, A.E., et al., *Smoking and multiple sclerosis: an updated meta-analysis.* PLoS One, 2011. **6**(1): p. e16149.
57. Friend, K.B., et al., *Smoking rates and smoking cessation among individuals with multiple sclerosis.* Disabil Rehabil, 2006. **28**(18): p. 1135-41.
58. Viglietta, V., et al., *Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis.* J Exp Med, 2004. **199**(7): p. 971-9.
59. Shi, Y., et al., *Critical regulation of CD4+ T cell survival and autoimmunity by beta-arrestin 1.* Nat Immunol, 2007. **8**(8): p. 817-24.
60. Trapp, B.D., et al., *Axonal transection in the lesions of multiple sclerosis.* N Engl J Med, 1998. **338**(5): p. 278-85.
61. Pérez-Cerdá, F., M.V. Sánchez-Gómez, and C. Matute, *The link of inflammation and neurodegeneration in progressive multiple sclerosis.* Multiple Sclerosis and Demyelinating Disorders, 2016. **1**(1): p. 9.
62. Bruck, W., *Inflammatory demyelination is not central to the pathogenesis of multiple sclerosis.* J Neurol, 2005. **252 Suppl 5**: p. v10-5.
63. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria.* Ann Neurol, 2011. **69**.
64. Filippi, M., et al., *MRI criteria for the diagnosis of multiple sclerosis: MAGNIMS consensus guidelines.* The Lancet Neurology, 2016. **15**(3): p. 292-303.
65. Polman, C.H., et al., *Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria.* Annals of Neurology, 2011. **69**(2): p. 292-302.
66. Pitt, D., et al., *Imaging cortical lesions in multiple sclerosis with ultra–high-field magnetic resonance imaging.* Archives of Neurology, 2010. **67**(7): p. 812-818.
67. Lycklama, G., et al., *Spinal-cord MRI in multiple sclerosis.* The Lancet Neurology. **2**(9): p. 555-562.
68. Klawiter, E.C., *Current and New Directions in MRI in Multiple Sclerosis.* Continuum : Lifelong Learning in Neurology, 2013. **19**(4 Multiple Sclerosis): p. 1058-1073.
69. Marcus, J.F. and E.L. Waubant, *Updates on Clinically Isolated Syndrome and Diagnostic Criteria for Multiple Sclerosis.* The Neurohospitalist, 2013. **3**(2): p. 65-80.
70. Compston, A. and A. Coles, *Multiple sclerosis.* Lancet, 2008. **372**(9648): p. 1502-17.
71. Piras, M.R., et al., *Longitudinal study of cognitive dysfunction in multiple sclerosis: neuropsychological, neuroradiological, and neurophysiological findings.* Journal of Neurology, Neurosurgery & Psychiatry, 2003. **74**(7): p. 878-885.

72. Pelosi, L., et al., *Working memory impairment in early multiple sclerosis. Evidence from an event-related potential study of patients with clinically isolated myelopathy.* Brain, 1997. **120 (Pt 11)**: p. 2039-58.
73. Ron, M.A., M.M. Callanan, and E.K. Warrington, *Cognitive abnormalities in multiple sclerosis: a psychometric and MRI study.* Psychol Med, 1991. **21**(1): p. 59-68.
74. Cobble, N., *The Rehabilitative Management of Patients with Multiple Sclerosis.* Journal of Neurologic Rehabilitation, 1992. **6**(3): p. 141-145.
75. Macniven, J.A., et al., *Stroop performance in multiple sclerosis: information processing, selective attention, or executive functioning?* J Int Neuropsychol Soc, 2008. **14**(5): p. 805-14.
76. Bergendal, G., S. Fredrikson, and O. Almkvist, *Selective decline in information processing in subgroups of multiple sclerosis: an 8-year longitudinal study.* Eur Neurol, 2007. **57**(4): p. 193-202.
77. Denney, D.R., L.A. Sworowski, and S.G. Lynch, *Cognitive impairment in three subtypes of multiple sclerosis.* Archives of Clinical Neuropsychology, 2005. **20**(8): p. 967-981.
78. Rao, S.M., et al., *Memory disturbance in chronic progressive multiple sclerosis.* Arch Neurol, 1984. **41**(6): p. 625-31.
79. Nocentini, U., et al., *Cognitive dysfunction in patients with relapsing-remitting multiple sclerosis.* Mult Scler, 2006. **12**(1): p. 77-87.
80. Arnett, P.A., F.H. Barwick, and J.E. Beeney, *Depression in multiple sclerosis: review and theoretical proposal.* J Int Neuropsychol Soc, 2008. **14**(5): p. 691-724.
81. Minden, S.L. and R.B. Schiffer, *Depression and mood disorders in multiple sclerosis.* Cognitive and Behavioral Neurology, 1991. **4**(1): p. 62-77.
82. Rabins, P.V., et al., *Structural brain correlates of emotional disorder in multiple sclerosis.* Brain, 1986. **109 (Pt 4)**: p. 585-97.
83. Bakshi, R., et al., *Brain MRI lesions and atrophy are related to depression in multiple sclerosis.* Neuroreport, 2000. **11**(6): p. 1153-8.
84. Sabatini, U., et al., *Involvement of the limbic system in multiple sclerosis patients with depressive disorders.* Biol Psychiatry, 1996. **39**(11): p. 970-5.
85. Baumstarck-Barrau, K., et al., *Cognitive function and quality of life in multiple sclerosis patients: a cross-sectional study.* BMC Neurol, 2011. **11**: p. 17.
86. Grima, D.T., et al., *Cost and health related quality of life consequences of multiple sclerosis.* Mult Scler, 2000. **6**(2): p. 91-8.
87. Wang, J.L., et al., *Major depression and quality of life in individuals with multiple sclerosis.* Int J Psychiatry Med, 2000. **30**(4): p. 309-17.
88. Henriksson, F., et al., *Costs, quality of life and disease severity in multiple sclerosis: a cross-sectional study in Sweden.* Eur J Neurol, 2001. **8**(1): p. 27-35.
89. *Burden of illness of multiple sclerosis: Part II: Quality of life. The Canadian Burden of Illness Study Group.* Can J Neurol Sci, 1998. **25**(1): p. 31-8.

90. Beatty, W.W., et al., *Demographic, clinical, and cognitive characteristics of multiple sclerosis patients who continue to work.* Journal of Neurologic Rehabilitation, 1995. **9**(3): p. 167-173.
91. Kornblith, A.B., N.G. La Rocca, and H.M. Baum, *Employment in individuals with multiple sclerosis.* International Journal of Rehabilitation Research, 1986. **9**(2): p. 155-166.
92. Rao, S.M., et al., *Cognitive dysfunction in multiple sclerosis. II. Impact on employment and social functioning.* Neurology, 1991. **41**(5): p. 692-696.
93. LaRocca, N., et al., *Factors associated with unemployment of patients with multiple sclerosis.* Journal of chronic diseases, 1985. **38**(2): p. 203-210.
94. Barak, Y. and A. Achiron, *Effect of interferon-beta-1b on cognitive functions in multiple sclerosis.* Eur Neurol, 2002. **47**(1): p. 11-4.
95. Patti, F., et al., *Subcutaneous Interferon Beta-1a Has a Positive Effect on Cognitive Performance in Mildly Disabled Patients with Relapsing-Remitting Multiple Sclerosis: 2-Year Results from the COGIMUS Study.* Therapeutic Advances in Neurological Disorders, 2009. **2**(2): p. 67-77.
96. Krupp, L.B., et al., *Donepezil improved memory in multiple sclerosis in a randomized clinical trial.* Neurology, 2004. **63**(9): p. 1579-85.
97. Christodoulou, C., et al., *Treatment of cognitive impairment in multiple sclerosis: is the use of acetylcholinesterase inhibitors a viable option?* CNS Drugs, 2008. **22**(2): p. 87-97.
98. Stuss, D.T., *The future of cognitive neurorehabilitation.* Neuropsychological rehabilitation, 2011. **21**(5): p. 755-768.
99. Chiaravalloti, N.D., H.M. Genova, and J. DeLuca, *Cognitive Rehabilitation in Multiple Sclerosis: The Role of Plasticity.* Frontiers in Neurology, 2015. **6**: p. 67.
100. das Nair, R., K.-J. Martin, and N.B. Lincoln, *Memory rehabilitation for people with multiple sclerosis.* Cochrane Database of Systematic Reviews, 2016(3).
101. Mitolo, M., et al., *Cognitive rehabilitation in multiple sclerosis: A systematic review.* Journal of the Neurological Sciences, 2015. **354**(1–2): p. 1-9.
102. Bever, C., et al., *The brief repeatable battery of neuropsychological tests for multiple sclerosis: a preliminary serial study.* Multiple Sclerosis, 1995. **1**(3): p. 165-169.
103. Langdon, D.W., et al., *Recommendations for a Brief International Cognitive Assessment for Multiple Sclerosis (BICAMS).* Mult Scler, 2012. **18**(6): p. 891-8.
104. Smith, A., *Symbol digit modalities test (SDMT) manual (revised) Western Psychological Services.* Los Angeles, 1982.
105. Buschke, H., *Selective reminding for analysis of memory and learning.* Journal of Verbal Learning and Verbal Behavior, 1973. **12**(5): p. 543-550.
106. Delis, D.C., et al., *CVLT-II: California verbal learning test: adult version*. 2000: Psychological Corporation.
107. Lerdal, A., et al., *A prospective study of patterns of fatigue in multiple sclerosis.* Eur J Neurol, 2007. **14**(12): p. 1338-43.

108. Ford, H., P. Trigwell, and M. Johnson, *The nature of fatigue in multiple sclerosis.* J Psychosom Res, 1998. **45**(1): p. 33-8.
109. Minden, S.L., et al., *The Sonya Slifka Longitudinal Multiple Sclerosis Study: methods and sample characteristics.* Mult Scler, 2006. **12**(1): p. 24-38.
110. Janardhan, V. and R. Bakshi, *Quality of life in patients with multiple sclerosis: the impact of fatigue and depression.* J Neurol Sci, 2002. **205**(1): p. 51-8.
111. Motl, R.W. and E. McAuley, *Symptom cluster as a predictor of physical activity in multiple sclerosis: preliminary evidence.* J Pain Symptom Manage, 2009. **38**(2): p. 270-80.
112. Putzki, N., et al., *Prevalence and severity of multiple-sclerosis-associated fatigue in treated and untreated patients.* Eur Neurol, 2008. **59**(3-4): p. 136-42.
113. Braley, T.J. and R.D. Chervin, *Fatigue in Multiple Sclerosis: Mechanisms, Evaluation, and Treatment.* Sleep, 2010. **33**(8): p. 1061-1067.
114. Kluger, B.M., L.B. Krupp, and R.M. Enoka, *Fatigue and fatigability in neurologic illnesses: proposal for a unified taxonomy.* Neurology, 2013. **80**(4): p. 409-16.
115. Guidelines, M.S.C.f.C.P., *Fatigue and multiple sclerosis: evidence-based management strategies for fatigue in multiple sclerosis*. 1998, Washington, DC: Paralyzed Veterans of America.
116. Krupp, L.B., et al., *The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus.* Archives of neurology, 1989. **46**(10): p. 1121-1123.
117. Motl, R.W., et al., *Symptomatic correlates of six-minute walk performance in persons with multiple sclerosis.* Eur J Phys Rehabil Med, 2013. **49**(1): p. 59-66.
118. Surakka, J., et al., *Effects of aerobic and strength exercise on motor fatigue in men and women with multiple sclerosis: a randomized controlled trial.* Clin Rehabil, 2004. **18**(7): p. 737-46.
119. Seamon, B.A. and M.O. Harris-Love, *Clinical Assessment of Fatigability in Multiple Sclerosis: A Shift from Perception to Performance.* Frontiers in Neurology, 2016. **7**: p. 194.
120. Geschwind, N., *DISCONNEXION SYNDROMES IN ANIMALS AND MAN1.* Brain, 1965. **88**(2): p. 237-237.
121. Catani, M. and D.H. ffytche, *The rises and falls of disconnection syndromes.* Brain, 2005. **128**(10): p. 2224-2239.
122. Calabrese, P. and I.K. Penner, *Cognitive dysfunctions in multiple sclerosis – a "multiple disconnection syndrome"?* Journal of Neurology, 2007. **254**(2): p. II18-II21.
123. Van Essen, D.C., et al., *The Human Connectome Project: a data acquisition perspective.* Neuroimage, 2012. **62**(4): p. 2222-31.
124. Pauling, L. and C.D. Coryell, *The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin.* Proceedings of the National Academy of Sciences of the United States of America, 1936. **22**(4): p. 210-216.
125. Ogawa, S., et al., *Brain magnetic resonance imaging with contrast dependent on blood oxygenation.* Proceedings of the National

Academy of Sciences of the United States of America, 1990. **87**(24): p. 9868-9872.
126. Mansfield, P., *Multi-planar image formation using NMR spin echoes.* Journal of Physics C: Solid State Physics, 1977. **10**(3): p. L55.
127. Belliveau, J.W., et al., *Magnetic Resonance Imaging Mapping of Brain Function Human Visual Cortex.* Investigative radiology, 1992. **27**(0 2): p. S59-S65.
128. Blamire, A.M., et al., *Dynamic mapping of the human visual cortex by high-speed magnetic resonance imaging.* Proc Natl Acad Sci U S A, 1992. **89**(22): p. 11069-73.
129. Glover, G.H., *Overview of Functional Magnetic Resonance Imaging.* Neurosurgery clinics of North America, 2011. **22**(2): p. 133-139.
130. Rovira, À., et al., *Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis-clinical implementation in the diagnostic process.* Nat Rev Neurol, 2015. **11**.
131. Biswal, B., et al., *Functional connectivity in the motor cortex of resting human brain using echo-planar MRI.* Magn Reson Med, 1995. **34**(4): p. 537-41.
132. Fox, M.D., et al., *The human brain is intrinsically organized into dynamic, anticorrelated functional networks.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(27): p. 9673-9678.
133. Sbardella, E., et al., *Resting-State fMRI in MS: General Concepts and Brief Overview of Its Application.* BioMed Research International, 2015. **2015**: p. 212693.
134. Basser, P.J., J. Mattiello, and D. LeBihan, *Estimation of the effective self-diffusion tensor from the NMR spin echo.* J Magn Reson B, 1994. **103**(3): p. 247-54.
135. Basser, P.J., et al., *In vivo fiber tractography using DT-MRI data.* Magn Reson Med, 2000. **44**(4): p. 625-32.
136. Pagani, E., et al., *A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis.* Neuroimage, 2005. **26**(1): p. 258-65.
137. Simon, J.H., et al., *Identification of fibers at risk for degeneration by diffusion tractography in patients at high risk for MS after a clinically isolated syndrome.* J Magn Reson Imaging, 2006. **24**(5): p. 983-8.
138. Alexander, A.L., et al., *Diffusion tensor imaging of the brain.* Neurotherapeutics, 2007. **4**(3): p. 316-329.
139. Srinivasan, R., et al., *Evidence of elevated glutamate in multiple sclerosis using magnetic resonance spectroscopy at 3 T.* Brain, 2005. **128**(Pt 5): p. 1016-25.
140. Helms, G., *Volume correction for edema in single-volume proton MR spectroscopy of contrast-enhancing multiple sclerosis lesions.* Magn Reson Med, 2001. **46**(2): p. 256-63.
141. Filippi, M. and F. Agosta, *Magnetization Transfer MRI in Multiple Sclerosis.* Journal of Neuroimaging, 2007. **17**: p. 22S-26S.
142. Heylighen, F., P. Cilliers, and C. Gerhenson, *Complexity and Philosophy*, in *Complexity, Science and Society*, J. Bogg and R. Geyer, Editors. 2007, Radcliffe Publishing: Oxford, UK.

143. Kesić, S., *Systems biology, emergence and antireductionism.* Saudi Journal of Biological Sciences, 2016. **23**(5): p. 584-591.
144. Ulanowicz, R.E., *The balance between adaptability and adaptation.* Biosystems, 2002. **64**(1-3): p. 13-22.
145. Chartrand, G. and P. Zhang, *A First Course in Graph Theory*. 2012, Mineola, NY: Dover Publications.
146. Euler, L., *Solutio problematis ad geometriam situs pertinentis.* Commentarii Academiae Scientarium Imperialis Petropolitanae, 1736. **8**: p. 128-140.
147. Moreno, J.L. and H.H. Jennings, *Who shall survive?* Vol. 58. 1934: JSTOR.
148. Erdos, P. and A. Rényi, *On the evolution of random graphs.* Publ. Math. Inst. Hung. Acad. Sci, 1960. **5**(1): p. 17-60.
149. Watts, D.J. and S.H. Strogatz, *Collective dynamics of 'small-world'networks.* Nature, 1998. **393**(6684): p. 440-442.
150. Albert, R. and A.-L. Barabási, *Statistical mechanics of complex networks.* Reviews of modern physics, 2002. **74**(1): p. 47.
151. Newman, M.E., *Modularity and community structure in networks.* Proceedings of the National Academy of Sciences, 2006. **103**(23): p. 8577-8582.
152. Staffen, W., et al., *Cognitive function and fMRI in patients with multiple sclerosis: evidence for compensatory cortical activation during an attention task.* Brain, 2002. **125**(6): p. 1275-1282.
153. Audoin, B., et al., *Compensatory cortical activation observed by fMRI during a cognitive task at the earliest stage of MS.* Hum Brain Mapp, 2003. **20**(2): p. 51-8.
154. Mainero, C., et al., *fMRI evidence of brain reorganization during attention and memory tasks in multiple sclerosis.* Neuroimage, 2004. **21**(3): p. 858-67.
155. Hulst, H.E., et al., *Functional adaptive changes within the hippocampal memory system of patients with multiple sclerosis.* Hum Brain Mapp, 2012. **33**(10): p. 2268-80.
156. Rocca, M.A., et al., *Functional correlates of cognitive dysfunction in multiple sclerosis: A multicenter fMRI Study.* Human Brain Mapping, 2014. **35**(12): p. 5799-5814.
157. Roosendaal, S.D., et al., *Resting state networks change in clinically isolated syndrome.* Brain, 2010. **133**(Pt 6): p. 1612-21.
158. Rocca, M.A., et al., *Default-mode network dysfunction and cognitive impairment in progressive MS.* Neurology, 2010. **74**(16): p. 1252-9.
159. Schoonheim, M.M., J.J. Geurts, and F. Barkhof, *The limits of functional reorganization in multiple sclerosis.* Neurology, 2010. **74**(16): p. 1246-7.
160. Bonavita, S., et al., *Distributed changes in default-mode resting-state connectivity in multiple sclerosis.* Mult Scler, 2011. **17**(4): p. 411-22.
161. Cruz-Gomez, A.J., et al., *The link between resting-state functional connectivity and cognition in MS patients.* Mult Scler, 2014. **20**(3): p. 338-48.
162. Leavitt, V.M., J. Paxton, and J.F. Sumowski, *Default network connectivity is linked to memory status in multiple sclerosis.* J Int Neuropsychol Soc, 2014. **20**(9): p. 937-44.

163. Louapre, C., et al., *Brain networks disconnection in early multiple sclerosis cognitive deficits: an anatomofunctional study.* Hum Brain Mapp, 2014. **35**(9): p. 4706-17.
164. Faivre, A., et al., *Assessing brain connectivity at rest is clinically relevant in early multiple sclerosis.* Mult Scler, 2012. **18**(9): p. 1251-8.
165. Zhou, F., et al., *Altered Inter-Subregion Connectivity of the Default Mode Network in Relapsing Remitting Multiple Sclerosis: A Functional and Structural Connectivity Study.* PLoS ONE, 2014. **9**(7): p. e101198.
166. Wojtowicz, M., et al., *Altered functional connectivity and performance variability in relapsing-remitting multiple sclerosis.* Mult Scler, 2014. **20**(11): p. 1453-63.
167. Hawellek, D.J., et al., *Increased functional connectivity indicates the severity of cognitive impairment in multiple sclerosis.* Proc Natl Acad Sci U S A, 2011. **108**(47): p. 19066-71.
168. Rocca, M.A., et al., *Posterior brain damage and cognitive impairment in pediatric multiple sclerosis.* Neurology, 2014. **82**(15): p. 1314-21.
169. Loitfelder, M., et al., *Abnormalities of Resting State Functional Connectivity Are Related to Sustained Attention Deficits in MS.* PLoS ONE, 2012. **7**(8): p. e42862.
170. Cader, S., et al., *Reduced brain functional reserve and altered functional connectivity in patients with multiple sclerosis.* Brain, 2006. **129**(Pt 2): p. 527-37.
171. Bonnet, M.C., et al., *Cognitive compensation failure in multiple sclerosis.* Neurology, 2010. **75**(14): p. 1241-8.
172. Au Duong, M.V., et al., *Altered functional connectivity related to white matter changes inside the working memory network at the very early stage of MS.* J Cereb Blood Flow Metab, 2005. **25**(10): p. 1245-53.
173. Ranjeva, J.-P., et al., *Structural and functional surrogates of cognitive impairment at the very early stage of multiple sclerosis.* Journal of the Neurological Sciences, 2006. **245**(1–2): p. 161-167.
174. Forn, C., et al., *Functional magnetic resonance imaging correlates of cognitive performance in patients with a clinically isolated syndrome suggestive of multiple sclerosis at presentation: an activation and connectivity study.* Mult Scler, 2012. **18**(2): p. 153-63.
175. Nava, E. and B. Röder, *Chapter 12 - Adaptation and maladaptation: insights from brain plasticity*, in *Progress in Brain Research*, C.E.C.J.F.K. Andrea M. Green and L. Franco, Editors. 2011, Elsevier. p. 177-194.
176. Sumowski, J.F., et al., *Intellectual enrichment is linked to cerebral efficiency in multiple sclerosis: functional magnetic resonance imaging evidence for cognitive reserve.* Brain, 2010. **133**(Pt 2): p. 362-74.
177. He, Y., et al., *Impaired small-world efficiency in structural cortical networks in multiple sclerosis associated with white matter lesion load.* Brain, 2009. **132**(12): p. 3366-3379.
178. Kocevar, G., et al., *Graph Theory-Based Brain Connectivity for Automatic Classification of Multiple Sclerosis Clinical Courses.* Frontiers in Neuroscience, 2016. **10**: p. 478.
179. Llufriu, S., et al., *Structural networks involved in attention and executive functions in multiple sclerosis.* NeuroImage : Clinical, 2017. **13**: p. 288-296.

180. Fleischer, V., et al., *Increased structural white and grey matter network connectivity compensates for functional decline in early multiple sclerosis.* Multiple Sclerosis Journal, 2016: p. 1352458516651503.
181. Li, Y., et al., *Diffusion Tensor Imaging Based Network Analysis Detects Alterations of Neuroconnectivity in Patients with Clinically Early Relapsing Remitting Multiple Sclerosis.* Human brain mapping, 2013. **34**(12): p. 3376-3391.
182. Shu, N., et al., *Diffusion Tensor Tractography Reveals Disrupted Topological Efficiency in White Matter Structural Networks in Multiple Sclerosis.* Cerebral Cortex, 2011. **21**(11): p. 2565-2577.
183. Liu, Y., et al., *Functional Brain Network Alterations in Clinically Isolated Syndrome and Multiple Sclerosis: A Graph-based Connectome Study.* Radiology, 2017. **282**(2): p. 534-541.
184. Rocca, M.A., et al., *Impaired functional integration in multiple sclerosis: a graph theory study.* Brain Structure and Function, 2016. **221**(1): p. 115-131.
185. Schoonheim, M.M., et al., *Gender-related differences in functional connectivity in multiple sclerosis.* Mult Scler, 2012. **18**(2): p. 164-73.
186. Shu, N., et al., *Disrupted topological organization of structural and functional brain connectomes in clinically isolated syndrome and multiple sclerosis.* Scientific Reports, 2016. **6**: p. 29383.
187. Gamboa, O.L., et al., *Working memory performance of early MS patients correlates inversely with modularity increases in resting state functional connectivity networks.* Neuroimage, 2014. **94**: p. 385-95.
188. Schoonheim, M., et al., *Changes in functional network centrality underlie cognitive dysfunction and physical disability in multiple sclerosis.* Mult Scler, 2013. **20**(8): p. 1058-1065.
189. Hardmeier, M., et al., *Cognitive dysfunction in early multiple sclerosis: altered centrality derived from resting-state functional connectivity using magneto-encephalography.* PLoS One, 2012. **7**(7): p. e42087.
190. Schoonheim, M.M., et al., *Functional connectivity changes in multiple sclerosis patients: a graph analytical study of MEG resting state data.* Hum Brain Mapp, 2013. **34**(1): p. 52-61.
191. Tewarie, P., et al., *Functional brain network analysis using minimum spanning trees in Multiple Sclerosis: an MEG source-space study.* Neuroimage, 2014. **88**: p. 308-18.
192. Tewarie, P., et al., *Disruption of structural and functional networks in long-standing multiple sclerosis.* Hum Brain Mapp, 2014. **35**(12): p. 5946-61.
193. Shu, N., et al., *Disrupted topological organization of structural and functional brain connectomes in clinically isolated syndrome and multiple sclerosis.* Sci Rep, 2016. **6**: p. 29383.
194. Rocca, M.A., et al., *Impaired functional integration in multiple sclerosis: a graph theory study.* Brain Struct Funct, 2016. **221**(1): p. 115-31.
195. Van Schependom, J., et al., *Graph theoretical analysis indicates cognitive impairment in MS stems from neural disconnection.* Neuroimage Clin, 2014. **4**: p. 403-10.
196. Schoonheim, M.M., et al., *Changes in functional network centrality underlie cognitive dysfunction and physical disability in multiple sclerosis.* Mult Scler, 2014. **20**(8): p. 1058-65.

197. Filippi, M., et al., *The contribution of MRI in assessing cognitive impairment in multiple sclerosis.* Neurology, 2010. **75**(23): p. 2121-2128.
198. Ge, Y., *Multiple Sclerosis: The Role of MR Imaging.* American Journal of Neuroradiology, 2006. **27**(6): p. 1165-1176.
199. Arnett, P.A., et al., *Relationship between frontal lobe lesions and Wisconsin Card Sorting Test performance in patients with multiple sclerosis.* Neurology, 1994. **44**(3 Pt 1): p. 420-5.
200. Charil, A., et al., *Statistical mapping analysis of lesion location and neurological disability in multiple sclerosis: application to 452 patient data sets.* Neuroimage, 2003. **19**(3): p. 532-44.
201. Foong, J., et al., *Executive function in multiple sclerosis. The role of frontal lobe pathology.* Brain, 1997. **120**(Pt 1): p. 15-26.
202. Rovaris, M., et al., *Relation between MR abnormalities and patterns of cognitive impairment in multiple sclerosis.* Neurology, 1998. **50**(6): p. 1601-8.
203. Swirsky-Sacchetti, T., et al., *Neuropsychological and structural brain lesions in multiple sclerosis: a regional analysis.* Neurology, 1992. **42**(7): p. 1291-5.
204. Tsolaki, M., et al., *Correlation of dementia, neuropsychological and MRI findings in multiple sclerosis.* Dementia, 1994. **5**(1): p. 48-52.
205. Iannucci, G., et al., *Correlation of multiple sclerosis measures derived from T2-weighted, T1-weighted, magnetization transfer, and diffusion tensor MR imaging.[see comment].* Ajnr: American Journal of Neuroradiology, 2001. **22**(8): p. 1462-7.
206. Ciccarelli, O., et al., *Investigation of MS normal-appearing brain using diffusion tensor MRI with clinical correlations.* Neurology, 2001. **56**(7): p. 926-933.
207. Filippi, M., et al., *Diffusion tensor magnetic resonance imaging in multiple sclerosis.* Neurology, 2001. **56**(3): p. 304-311.
208. Guo, C.C., et al., *One-year test–retest reliability of intrinsic connectivity network fMRI in older adults.* Neuroimage, 2012. **61**(4): p. 1471-1483.
209. Werring, D.J., et al., *Diffusion tensor imaging of lesions and normal-appearing white matter in multiple sclerosis.* Neurology, 1999. **52**: p. 1626-1632.
210. Smith, S.M., et al., *Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data.* Neuroimage, 2006. **31**(4): p. 1487-505.
211. Blaschek, A., et al., *Early white matter changes in childhood multiple sclerosis: a diffusion tensor imaging study.* American Journal of Neuroradiology, 2013. **34**(10): p. 2015-2020.
212. Bodini, B., et al., *Exploring the relationship between white matter and gray matter damage in early primary progressive multiple sclerosis: an in vivo study with TBSS and VBM.* Human brain mapping, 2009. **30**(9): p. 2852-2861.
213. Cader, S., et al., *Discordant white matter N-acetylasparate and diffusion MRI measures suggest that chronic metabolic dysfunction contributes to axonal pathology in multiple sclerosis.* NeuroImage, 2007. **36**(1): p. 19.

214. Dineen, R.A., et al., *Extra-Hippocampal Subcortical Limbic Involvement Predicts Episodic Recall Performance in Multiple Sclerosis.* PloS one, 2012. **7**(10): p. e44942.
215. Giorgio, A., et al., *Relationships of brain white matter microstructure with clinical and MR measures in relapsing-remitting multiple sclerosis.* Journal of Magnetic Resonance Imaging, 2010. **31**(2): p. 309-316.
216. Hulst, H.E., et al., *Cognitive impairment in MS Impact of white matter integrity, gray matter volume, and lesions.* Neurology, 2013. **80**(11): p. 1025-1032.
217. Kern, K.C., et al., *Corpus callosal diffusivity predicts motor impairment in relapsing–remitting multiple sclerosis: a TBSS and tractography study.* Neuroimage, 2011. **55**(3): p. 1169-1177.
218. Liu, Y., et al., *Whole brain white matter changes revealed by multiple diffusion metrics in multiple sclerosis: a TBSS study.* European journal of radiology, 2012. **81**(10): p. 2826-2832.
219. Llufriu, S., et al., *Cognitive functions in multiple sclerosis: impact of gray matter integrity.* Multiple Sclerosis, 2013.
220. Mazerolle, E.L., et al., *Intra-individual variability in information processing speed reflects white matter microstructure in multiple sclerosis.* NeuroImage: clinical, 2013. **2**: p. 894-902.
221. Onu, M., et al., *Diffusion abnormality maps in demyelinating disease: correlations with clinical scores.* European journal of radiology, 2012. **81**(3): p. e386-e391.
222. Raz, E., et al., *Clinically isolated syndrome suggestive of multiple sclerosis: voxelwise regional investigation of white and gray matter.* Radiology, 2010. **254**(1): p. 227.
223. Roosendaal, S.D., et al., *Regional DTI differences in multiple sclerosis patients.* Neuroimage, 2009. **44**(4): p. 1397-1403.
224. Roosendaal, S.D., et al., *Resting state networks change in clinically isolated syndrome.* Brain, 2010. **133**(6): p. 1612-1621.
225. Sbardella, E., et al., *Assessing the correlation between grey and white matter damage with motor and cognitive impairment in multiple sclerosis patients.* PloS one, 2013. **8**(5): p. e63250.
226. Schoonheim, M.M., et al., *Sex-specific extent and severity of white matter damage in multiple sclerosis: Implications for cognitive decline.* Human brain mapping, 2013.
227. Yu, H.J., et al., *Multiple white matter tract abnormalities underlie cognitive impairment in RRMS.* Neuroimage, 2012. **59**(4): p. 3713-22.
228. Mori, S. and P.C. van Zijl, *Fiber tracking: principles and strategies - a technical review.* NMR Biomed, 2002. **15**(7-8): p. 468-80.
229. Falagas, M.E., et al., *Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses.* FASEB J, 2008. **22**(2): p. 338-42.
230. Kmet, L., R. Lee, and L. Cook, *Standard quality assessment criteria for evaluating primary research papers from a variety of fields*. Vol. 13. 2004, Edmonton, Alberta: Alberta Heritage Foundation for Medical Research.
231. Radua, J., et al., *A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps.* Eur Psychiatry, 2012. **17**(8): p. 605-11.

232. Hedges, L.V. and I. Olkin, *Statistical Methods for Meta-Analysis*. 1985, Orlando, FL: Academic Press.
233. Peters, B.D., et al., *White Matter Development in Adolescence: Diffusion Tensor Imaging and Meta-Analytic Results.* Schizophrenia Bulletin, 2012. **38**(6): p. 1308-1317.
234. Tench, C.R., et al., *Coordinate based meta-analysis of functional neuroimaging data; false discovery control and diagnostics.* PLoS One, 2013. **8**(7): p. e70143.
235. Simpson, S., et al., *Latitude is significantly associated with the prevalence of multiple sclerosis: a meta-analysis.* Journal of Neurology, Neurosurgery & Psychiatry, 2011. **82**(10): p. 1132-1141.
236. Lin, F., et al., *Diffusion tensor tractography-based group mapping of the pyramidal tract in relapsing-remitting multiple sclerosis patients.* AJNR Am J Neuroradiol, 2007. **28**(2): p. 278-82.
237. Zhang, Y., et al., *Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease.* Neurology, 2007. **68**(1): p. 13-19.
238. Takahashi, M., et al., *White matter microstructure of the cingulum and cerebellar peduncle is related to sustained attention and working memory: a diffusion tensor imaging study.* Neurosci Lett, 2010. **477**(2): p. 72-6.
239. Hasan, K.M., et al., *Diffusion tensor fractional anisotropy of the normal-appearing seven segments of the corpus callosum in healthy adults and relapsing-remitting multiple sclerosis patients.* J Magn Reson Imaging, 2005. **21**(6): p. 735-43.
240. Mesaros, S., et al., *Corpus callosum damage and cognitive dysfunction in benign MS.* Hum Brain Mapp, 2009. **30**(8): p. 2656-66.
241. Sharp, D.J., G. Scott, and R. Leech, *Network dysfunction after traumatic brain injury.* Nat Rev Neurol, 2014. **10**(3): p. 156-66.
242. Dogonowski, A.M., et al., *Multiple sclerosis impairs regional functional connectivity in the cerebellum.* Neuroimage Clin, 2014. **4**: p. 130-8.
243. He, Y., et al., *Impaired small-world efficiency in structural cortical networks in multiple sclerosis associated with white matter lesion load.* Brain, 2009. **132**(Pt 12): p. 3366-79.
244. Rocca, M., et al., *Default-mode network dysfunction and cognitive impairment in progressive MS.* Neurology, 2010. **74**(16): p. 1252-1259.
245. Van Schependom, J., et al., *Graph theoretical analysis indicates cognitive impairment in MS stems from neural disconnection.* NeuroImage: Clinical, 2014. **4**: p. 403-410.
246. Ni, H., et al., *Effects of number of diffusion gradient directions on derived diffusion tensor imaging indices in human brain.* AJNR Am J Neuroradiol, 2006. **27**(8): p. 1776-81.
247. Gobbi, C., et al., *Forceps minor damage and co-occurrence of depression and fatigue in multiple sclerosis.* Multiple Sclerosis Journal, 2014. **20**(12): p. 1633-1640.
248. Shen, Y., et al., *Depressive Symptoms in Multiple Sclerosis from an In Vivo Study with TBSS.* BioMed Research International, 2014. **2014**: p. 148465.

249. Zimmermann, H., et al., *Putaminal alteration in multiple sclerosis patients with spinal cord lesions.* Journal of Neural Transmission, 2015. **122**(10): p. 1465-1473.
250. Asaf, A., S. Evan, and A. Anat, *Injury to white matter tracts in relapsing–remitting multiple sclerosis: A possible therapeutic window within the first 5 years from onset using diffusion-tensor imaging tract-based spatial statistics.* NeuroImage : Clinical, 2015. **8**: p. 261-266.
251. Akbar, N., et al., *Alterations in Functional and Structural Connectivity in Pediatric-Onset Multiple Sclerosis.* PLoS ONE, 2016. **11**(1): p. e0145906.
252. Meijer, K.A., et al., *White matter tract abnormalities are associated with cognitive dysfunction in secondary progressive multiple sclerosis.* Multiple Sclerosis Journal, 2016. **22**(11): p. 1429-1437.
253. Welton, T., et al., *Reproducibility of Graph-Theoretic Brain Network Metrics: A Systematic Review.* Brain Connect, 2014. **5**(4): p. 193-202.
254. Petrella, J.R., *Use of graph theory to evaluate brain networks: a clinical tool for a small world?* Radiology, 2011. **259**(2): p. 317-20.
255. Quraan, M.A., et al., *Altered resting state brain dynamics in temporal lobe epilepsy can be observed in spectral power, functional connectivity and graph theory metrics.* PLoS One, 2013. **8**(7): p. e68609.
256. Andreotti, J., et al., *Repeatability analysis of global and local metrics of brain structural networks.* Brain Connect, 2014. **4**(3): p. 203-20.
257. Zuo, X.-N. and X.-X. Xing, *Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective.* Neuroscience & Biobehavioral Reviews, 2014. **45**: p. 100-118.
258. Shrout, P.E. and J.L. Fleiss, *Intraclass correlations: uses in assessing rater reliability.* Psychological bulletin, 1979. **86**(2): p. 420.
259. Bartlett, J. and C. Frost, *Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables.* Ultrasound in Obstetrics & Gynecology, 2008. **31**(4): p. 466-475.
260. McGraw, K.O. and S.P. Wong, *Forming inferences about some intraclass correlation coefficients.* Psychological methods, 1996. **1**(1): p. 30.
261. Faria, A.V., et al., *Atlas-based analysis of resting-state functional connectivity: Evaluation for reproducibility and multi-modal anatomy–function correlation studies.* Neuroimage, 2012. **61**(3): p. 613-621.
262. Jorge-Hernandez, F. and B. Garcia-Zapirain, *Reliability Test for Processing of Magnetic Resonance Images in Resting State Using Graph Theory.* Journal of Medical Imaging and Health Informatics, 2016. **6**(5): p. 1288-1292.
263. Kuntzelman, K. and V. Miskovic, *Reliability of graph metrics derived from resting-state human EEG.* Psychophysiology, 2017. **54**(1): p. 51-61.
264. Aurich, N.K., et al., *Evaluating the reliability of different preprocessing steps to estimate graph theoretical measures in resting state fMRI data.* Frontiers in Neuroscience, 2015. **9**: p. 48.

265. Bonilha, L., et al., *Reproducibility of the Structural Brain Connectome Derived from Diffusion Tensor Imaging.* PLoS ONE, 2015. **10**(9): p. e0135247.
266. Kuhn, T., et al., *Test-Retest Reliability of High Angular Resolution Diffusion Imaging Acquisition within Medial Temporal Lobe Connections Assessed via Tract Based Spatial Statistics, Probabilistic Tractography and a Novel Graph Theory Metric.* Brain imaging and behavior, 2016. **10**(2): p. 533-547.
267. Wang, J., et al., *Test-retest reliability of functional connectivity networks during naturalistic fMRI paradigms.* Hum Brain Mapp, 2017. **38**(4): p. 2226-2241.
268. Braun, U., et al., *Test–retest reliability of resting-state connectivity network characteristics using fMRI and graph theoretical measures.* Neuroimage, 2012. **59**(2): p. 1404-1412.
269. Deuker, L., et al., *Reproducibility of graph metrics of human brain functional networks.* Neuroimage, 2009. **47**(4): p. 1460-1468.
270. Liao, X.-H., et al., *Functional Brain Hubs and Their Test-Retest Reliability: A Multiband Resting-State Functional MRI Study.* NeuroImage, 2013. **83**: p. 969-982.
271. Niu, H., et al., *Test-retest reliability of graph metrics in functional brain networks: a resting-state fNIRS study.* PloS one, 2013. **8**(9): p. e72425.
272. Park, H.-J. and K. Friston, *Structural and functional brain networks: from connections to cognition.* Science, 2013. **342**(6158): p. 1238411.
273. Telesford, Q.K., et al., *Reproducibility of graph metrics in fMRI networks.* Frontiers in NEUROINFORMATICS, 2010. **4**: p. 117.
274. Wang, J.-H., et al., *Graph theoretical analysis of functional brain networks: test-retest evaluation on short-and long-term resting-state functional MRI data.* PloS one, 2011. **6**(7): p. e21976.
275. Weber, M.J., et al., *Reproducibility of functional network metrics and network structure: A comparison of task-related BOLD, resting ASL with BOLD contrast, and resting cerebral blood flow.* Cognitive, Affective, & Behavioral Neuroscience, 2013. **13**(3): p. 627-640.
276. Du, H.-X., et al., *Test–Retest Reliability of Graph Metrics in High-resolution Functional Connectomics: A Resting-State Functional MRI Study.* CNS Neuroscience & Therapeutics, 2015. **21**(10): p. 802-816.
277. Cao, H., et al., *Test–retest reliability of fMRI-based graph theoretical properties during working memory, emotion processing, and resting state.* NeuroImage, 2014. **84**: p. 888-900.
278. Jin, S.-H., et al., *How reliable are the functional connectivity networks of MEG in resting states?* Journal of neurophysiology, 2011. **106**(6): p. 2888-2895.
279. Schwarz, A.J. and J. McGonigle, *Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data.* Neuroimage, 2011. **55**(3): p. 1132-1146.
280. Termenon, M., et al., *Reliability of graph analysis of resting state fMRI using test-retest dataset from the Human Connectome Project.* NeuroImage, 2016. **142**: p. 172-187.
281. Liang, X., et al., *Effects of different correlation metrics and preprocessing factors on small-world brain functional networks: a resting-state functional MRI study.* PloS one, 2012. **7**(3): p. e32766.

282. Fan, Y., et al., *Test-retest reliability of the graph metrics in different rest conditions and sampling rates.* Annual Meeting of the Organization for Human Brain Mapping, 2012. **8**(9): p. e72425.
283. Buchanan, C.R., et al., *Test–retest reliability of structural brain networks from diffusion MRI.* NeuroImage, 2014. **86**: p. 231-243.
284. Cheng, H., et al., *Characteristics and variability of structural networks derived from diffusion tensor imaging.* Neuroimage, 2012. **61**(4): p. 1153-1164.
285. Duda, J.T., P.A. Cook, and J.C. Gee, *Reproducibility of graph metrics of human brain structural networks.* Frontiers in neuroinformatics, 2014. **8**.
286. Owen, J.P., et al., *Test–Retest Reliability of Computational Network Measurements Derived from the Structural Connectome of the Human Brain.* Brain connectivity, 2013. **3**(2): p. 160-176.
287. Parker, C., C. Clark, and J. Clayden, *Reproducibility of Whole-brain Structural Connectivity Networks.* Neuroimage, 2012. **61**: p. 1153-1164.
288. Vaessen, M., et al., *The effect and reproducibility of different clinical DTI gradient sets on small world brain connectivity measures.* Neuroimage, 2010. **51**(3): p. 1106-1116.
289. Bassett, D.S., et al., *Conserved and variable architecture of human white matter connectivity.* Neuroimage, 2011. **54**(2): p. 1262-1279.
290. Dennis, E.L., et al., *Test-retest reliability of graph theory measures of structural brain connectivity*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. 2012, Springer. p. 305-312.
291. Zhao, T., et al., *Test-retest reliability of white matter structural brain networks: a multiband diffusion MRI study.* Frontiers in Human Neuroscience, 2015. **9**: p. 59.
292. Biswal, B., et al., *Functional connectivity in the motor cortex of resting human brain using echo‐planar mri.* Magnetic resonance in medicine, 1995. **34**(4): p. 537-541.
293. Wang, J., et al., *Onset and Progression of With-the-Rule Astigmatism in Children with Infantile Nystagmus Syndrome.* Investigative Ophthalmology & Visual Science, 2010. **51**(1): p. 594-601.
294. Müller, R. and P. Büttner, *A critical discussion of intraclass correlation coefficients.* Statistics in medicine, 1994. **13**(23-24): p. 2465-2476.
295. Park, B., et al., *Are brain networks stable during a 24-hour period?* Neuroimage, 2012. **59**(1): p. 456-466.
296. Jenkinson, M., et al., *FSL.* Neuroimage, 2012. **62**(2): p. 782-90.
297. Desikan, R.S., et al., *An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.* NeuroImage, 2006. **31**(3): p. 968-980.
298. Destrieux, C., et al., *Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature.* Neuroimage, 2010. **53**(1): p. 1-15.
299. Friedman, L., G.H. Glover, and F. Consortium, *Reducing interscanner variability of activation in a multicenter fMRI study: controlling for signal-to-fluctuation-noise-ratio (SFNR) differences.* Neuroimage, 2006. **33**(2): p. 471-481.

300. Birn, R.M., et al., *The effect of scan length on the reliability of resting-state fMRI connectivity estimates.* Neuroimage, 2013. **83**: p. 550-558.
301. Forsyth, J.K., et al., *Reliability of functional magnetic resonance imaging activation during working memory in a multi-site study: Analysis from the North American Prodrome Longitudinal Study.* NeuroImage, 2014. **97**: p. 41-52.
302. Bennett, C.M. and M.B. Miller, *How reliable are the results from functional magnetic resonance imaging?* Annals of the New York Academy of Sciences, 2010. **1191**(1): p. 133-155.
303. Turner, R., et al., *Functional mapping of the human visual cortex at 4 and 1.5 tesla using deoxygenation contrast EPI.* Magn Reson Med, 1993. **29**(2): p. 277-9.
304. Kaza, E., U. Klose, and M. Lotze, *Comparison of a 32-channel with a 12-channel head coil: Are there relevant improvements for functional imaging?* Journal of Magnetic Resonance Imaging, 2011. **34**(1): p. 173-183.
305. Baumstarck, K., et al., *Measuring the Quality of Life in Patients with Multiple Sclerosis in Clinical Practice: A Necessary Challenge.* Multiple Sclerosis International, 2013. **2013**: p. 524894.
306. Deco, G., et al., *Rethinking segregation and integration: contributions of whole-brain modelling.* Nat Rev Neurosci, 2015. **16**(7): p. 430-439.
307. Omar, H., *Assessing the Effect of Midazolam on the Functional Connectivity using R-Graph Analysis*, in *Radiological and Imaging Sciences*. 2012, University of Nottingham.
308. Gorman, M.P., et al., *INcreased relapse rate in pediatric-onset compared with adult-onset multiple sclerosis.* Archives of Neurology, 2009. **66**(1): p. 54-59.
309. Lincoln, N.B., et al., *Cognitive Rehabilitation for Attention and Memory in people with Multiple Sclerosis: study protocol for a randomised controlled trial (CRAMMS).* Trials, 2015. **16**: p. 556.
310. Leavitt, V.M., et al., *Altered effective connectivity during performance of an information processing speed task in multiple sclerosis.* Mult Scler, 2012. **18**(4): p. 409-17.
311. Fischer, J.S., et al., *The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. National MS Society Clinical Outcomes Assessment Task Force.* Mult Scler, 1999. **5**(4): p. 244-50.
312. Gronwall, D.M.A., *Paced auditory serial-addition task: A measure of recovery from concussion.* Perceptual and motor skills, 1977. **44**: p. 367-373.
313. Beck, A.T., et al., *Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients.* J Pers Assess, 1996. **67**(3): p. 588-97.
314. Sadovnick, A.D., et al., *Depression and multiple sclerosis.* Neurology, 1996. **46**(3): p. 628-32.
315. Siegert, R. and D. Abernethy, *Depression in multiple sclerosis: a review.* Journal of Neurology, Neurosurgery, and Psychiatry, 2005. **76**(4): p. 469-475.
316. Fisk, J.D., et al., *The impact of fatigue on patients with multiple sclerosis.* Can J Neurol Sci, 1994. **21**(1): p. 9-14.

317. Buysse, D.J., et al., *The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.* Psychiatry Res, 1989. **28**(2): p. 193-213.
318. Attarian, H.P., et al., *The relationship of sleep disturbances and fatigue in multiple sclerosis.* Archives of Neurology, 2004. **61**(4): p. 525-528.
319. Bøe Lunde, H.M., et al., *Poor Sleep in Patients with Multiple Sclerosis.* PLoS ONE, 2012. **7**(11): p. e49996.
320. Fan, J., et al., *Testing the efficiency and independence of attentional networks.* J Cogn Neurosci, 2002. **14**(3): p. 340-7.
321. Smith, S.M., *Fast robust automated brain extraction.* Hum Brain Mapp, 2002. **17**(3): p. 143-55.
322. Zhang, Y., M. Brady, and S. Smith, *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.* IEEE Trans Med Imaging, 2001. **20**(1): p. 45-57.
323. Fischl, B., *FreeSurfer.* NeuroImage, 2012. **62**(2): p. 774-781.
324. Desikan, R.S., et al., *An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest.* Neuroimage, 2006. **31**(3): p. 968-80.
325. Bright, M.G. and K. Murphy, *Is fMRI "noise" really noise? Resting state nuisance regressors remove variance with network structure.* NeuroImage, 2015. **114**: p. 158-169.
326. Behrens, T.E., et al., *Characterization and propagation of uncertainty in diffusion-weighted MR imaging.* Magn Reson Med, 2003. **50**(5): p. 1077-88.
327. Amtmann, D., et al., *Comparison of the psychometric properties of two fatigue scales in multiple sclerosis.* Rehabilitation psychology, 2012. **57**(2): p. 159-166.
328. Crawford, P. and N.J. Webster, *Assessment of Depression in Multiple Sclerosis.* International Journal of MS Care, 2009. **11**(4): p. 167-173.
329. Drake, A.S., et al., *Psychometrics and normative data for the Multiple Sclerosis Functional Composite: replacing the PASAT with the Symbol Digit Modalities Test.* Mult Scler, 2010. **16**(2): p. 228-37.
330. Duyn, J., *Spontaneous fMRI activity during resting wakefulness and sleep.* Prog Brain Res, 2011. **193**: p. 295-305.
331. Patriat, R., et al., *The effect of resting condition on resting-state fMRI reliability and consistency: a comparison between resting with eyes open, closed, and fixated.* Neuroimage, 2013. **78**: p. 463-73.
332. Friedman, L. and G.H. Glover, *Report on a multicenter fMRI quality assurance protocol.* Journal of Magnetic Resonance Imaging, 2006. **23**(6): p. 827-839.
333. Watts, D.J. and S.H. Strogatz, *Collective dynamics of /`small-world/' networks.* Nature, 1998. **393**(6684): p. 440-442.
334. Bassett, D.S. and E. Bullmore, *Small-world brain networks.* The neuroscientist, 2006. **12**(6): p. 512-523.
335. Fornito, A., A. Zalesky, and E.T. Bullmore, *Network scaling effects in graph analytic studies of human resting-state FMRI data.* Frontiers in systems neuroscience, 2010. **4**.
336. Hosseini, S.M., F. Hoeft, and S.R. Kesler, *GAT: a graph-theoretical analysis toolbox for analyzing between-group differences in large-scale*

structural and functional brain networks. PLoS One, 2012. **7**(7): p. e40709.
337. Rubinov, M. and O. Sporns, *Complex network measures of brain connectivity: uses and interpretations.* Neuroimage, 2010. **52**(3): p. 1059-69.
338. Welton, T., et al., *Reproducibility of graph-theoretic brain network metrics: a systematic review.* Brain Connectivity, 2015. **5**(4): p. 193-202.
339. Austin, P.C. and E.W. Steyerberg, *The number of subjects per variable required in linear regression analyses.* Journal of Clinical Epidemiology, 2015. **68**(6): p. 627-636.
340. Raichle, M.E., *The Restless Brain.* Brain Connectivity, 2011. **1**(1): p. 3-12.
341. Hlinka, J., D. Hartman, and M. Palus, *Small-world topology of functional connectivity in randomly connected dynamical systems.* Chaos, 2012. **22**(3): p. 033107.
342. Tun, P.A. and M.E. Lachman, *Age Differences in Reaction time and Attention in a National Telephone Sample of Adults: Education, Sex, and Task Complexity Matter.* Developmental psychology, 2008. **44**(5): p. 1421-1429.
343. Hutchison, R.M., et al., *Dynamic functional connectivity: Promise, issues, and interpretations.* NeuroImage, 2013. **80**: p. 10.1016/j.neuroimage.2013.05.079.
344. Sakoglu, U., et al., *A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia.* MAGMA, 2010. **23**(5-6): p. 351-66.
345. Shirer, W.R., et al., *Decoding subject-driven cognitive states with whole-brain connectivity patterns.* Cereb Cortex, 2012. **22**(1): p. 158-65.
346. Hutchison, R.M., et al., *Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques.* Hum Brain Mapp, 2013. **34**(9): p. 2154-77.
347. Box, G.E. and N.R. Draper, *Empirical model-building and response surfaces*. Vol. 424. 1987: Wiley New York.