THE UNIVERSITY
of EDINBURGH

A longitudinal study of cognitive changes in MS – dimensionality, predictors and self-perception of change

Ramune Dirvanskiene

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Edinburgh

2016

Declaration

I declare that this thesis is my own composition, and that the materials contained in it describe my own work. It has not been submitted for any other degree of professional qualification. All sources of information have been acknowledged.

Ramune Dirvanskiene

Abstract

Background: Multiple sclerosis (MS) is a neurological disorder and the most frequent neurological cause of disability in young adults. 40-65% of MS patients experience cognitive difficulties (Bennedict et al., 2012), with problems in memory, attention and information processing speed being most frequently reported. However, visuoperceptual and language functions are much less studied in MS, but the few studies that looked into them (Vleugels, 2001; Grossman, 1995) have found prevalence of significant posterior deficits in MS. Up to today no study has investigated the domain-specificity of cognitive dysfunction in MS and its longitudinal progression.

Aims: The primary aim of this project was to investigate the dimensionality of MS-related cognitive impairment longitudinally. The second aim was to determine the predictors of the observed longitudinal changes. The third aim was to investigate whether the participants themselves were aware of their cognitive changes, and what predicted the self-perception of change.

Methods: To address these aims I followed a sample of MS patients and compared their performance on cognitive tests measuring five cognitive domains (verbal memory, visuospatial memory, processing speed, visuoperceptual and language) at baseline and at follow-up three years apart. Then I've composed separate models to explain the predictors first of the actual changes, and then of the perceived changes in performance. Moreover, as part of this project I have analysed pre-existing data to evaluate the instruments and optimized the baseline test battery for use in performing the follow-up assessments.

Findings: I have managed to collect follow-up data on 82 MS patients and 23 matching healthy controls, acquiring high (76% and 79% respectively) recruitment rates. My MS sample (24% PPMS, 34% SPMS and 46% RRMS) was representable of the overall MS population. I found that deficits were seen in all cognitive domains (none were spared) and that new deficits were picked up sporadically, although with higher predisposition towards the information processing speed, visuoperceptual and memory domains. The new deficits showed the tendency to slowly accumulate,

leading to development of major problems with longer disease duration. Interestingly it was found that even though the factors that influenced cognitive decline were specific for each of the cognitive domains, however, neurological disability, MS type and levels of depression were the most common predictors of change in cognitive functioning. I found that in general MS patients perceived longitudinal changes on the BRBN battery more accurately than on visuoperceptual and language tests, and the factors that played a role in the self-perception of change were executive dysfunction, neurological disability and MS impact.

Implications: The results of this study add significant contribution to the field of longitudinal change in cognition in MS. Not only I explored the dimensionality of MS-related cognitive deficits, but also examined the factors that led to poorer performance, and the patients' own perspective of their cognitive change. Moreover, with this project I have addressed common problems in the field of longitudinal research in MS – definition of normal variation in performance; the sensitivity of cognitive tests to pick up MS-related deficits; and heterogeneity of cognitive impairments in MS; - and I have used the performance of my own controls in attempts to account for all of that. I believe that this study will be of interest not only to those who specialize in cognitive functioning in MS, but also to those who question the methods employed in clinical research to define impairments and to account for individual differences.

Acknowledgements

I would like to acknowledge all the many people who have helped me through this journey and without whose support this would not have been possible.

Firstly my academic supervisors, Thomas Bak, Sharon Abrahams and Peter Connick. Through their joint supervision I have been fortunate to have had a wonderful mix of expertise, guidance, and support. They have helped me identify the areas that I need to work upon, both theoretical and methodological, and without their help I wouldn't have arrived to completing this project. I would also like to thank all the people at the Anne Rowling Regenerative Neurology Clinic for all their help, especially huge thanks to Siddharthan Chandran, Shuna Colville, and Mara Sittampalam, without whose help this project would not have even started.

For their moral, emotional, and financial support when needed, the highest gratitude is expressed to my parents, who have never stopped believing in me. Thank you to my husband for being my rock, and sometimes my shoulder to cry on; and for all of the chocolate bars and ice-cream that you had to bring home after work.

A final thanks to all of the participants in this study, without whose cooperation this research would not have been possible. In particular the generosity of MS patients and their families, who motivated me to continue with research, and who by sharing their stories gave inspiration for my future projects.

This thesis is dedicated to my husband Simonas. Thank you for your patience.

Contents

List of tables

List of figures

Chapter One. Introduction - literature review and implications for the current study

1.1. Multiple Sclerosis

1.1.1. MS definition, causes, prevalence, life expectancy, and cost

I. MS definition

Multiple sclerosis (MS) is a chronic neurodegenerative disorder, caused mainly by autoimmune destruction of the myelin sheath around the neurons in the central nervous system. It is the most frequent demyelinating disease and a leading cause for permanent disability in young adults. MS is characterized by a variable disease course that causes problems with bodily functions, movement, and cognition. Most often MS symptoms follow a pattern of relapses with near-complete recovery, followed by gradual progressive accumulation of neurological deficits.

II. MS causes

The cause of MS is currently unknown, and is likely associated with genetics as well as multiple environmental factors, such as geographical latitude, sunlight exposure, vitamin D deficiency and certain viruses (Milo & Kahana, 2010). None of the causes by themselves appear to be sufficient for the development of MS, and MS is likely to be caused by the complex interaction of all of these factors (Pryse-Phillips, 2001).

III. MS prevalence and life expectancy

It is estimated that 2.5 million people globally have MS with a prevalence of 30 in 100 000 people; and Scotland is among the countries with the highest prevalence in the world with 188 in 100 000 people in population (Kingwell et al., 2013). MS is 2.4 times more common in women than in men, with the prevalence reaching 285.5 per 100 000 in women and 113.1 per 100 000 in men in UK (Mackenzie, Morant, Bloomfield, MacDonald, & O'riordan, 2014). This gender discrepancy in

prevalence is greatest in patients with relapsing-remitting type of MS, and almost non-existent in primary progressive type (Noseworthy et al., 2000).

MS is typically diagnosed between ages 20 and 60 with peak incidence occurring around age 40 (Mackenzie, et al., 2014). The median survival is 35 to 42 years after diagnosis (Poser, Kurtzke, Poser, & Schlaf, 1989), leading patients with MS (pwMS) to have six to seven years shorter life expectancy than the general population (Sadovnick, Ebers, Wilson, & Paty, 1992).

IV. MS cost

MS has the tendency to strike during the prime employment years and is associated with great costs to the individual and society, ranging from 8 528 to 54 244 USD per patient per year in medical, non-medical and indirect costs (Adelman, Rane, & Villa, 2013), such as the patient having to medically retire and the partner (or parent or child) having to take time off work to take care of the patient.

1.1.2. Diagnosis of MS – pathology and symptoms

Currently in clinical practice MS is diagnosed following the guidelines of the revised McDonald Criteria (Polman et al., 2011) that involves clinical examination of MS symptoms and MS pathology.

I. MS pathology

The examination of MS pathology consists of the analysis of Magnetic Resonance Imaging (MRI) and cerebrospinal fluid data.

On MRI, pwMS exhibit multifocal lesions, most often in the white matter of periventricular, brain stem, cerebellum, and spinal cord areas. The distribution, as well as the rate of acquiring the lesions is very heterogeneous. On occasions when there is diagnostic uncertainty about the rate of MS activity, repeated MRI after several months can provide evidence that MS is active and the lesions are disseminated in time (Noseworthy et al., 2000). While MS is traditionally considered an inflammatory, white matter disease, degeneration of gray matter is increasingly recognized as a primary contributor to progressive cognitive decline (Trapp & Nave, 2008), and can occur independently (Frischer et al., 2009)

In MS diagnostics cerebrospinal fluid analysis is used to pick up increased intrathecal synthesis of immunoglobulins with moderate lymphocytic pleocytosis (Noseworthy et al., 2000), that indicate presence of auto-immune mechanisms. Presence of auto-immune reactions are considered to be the evidence of the immune system attacks on the nerve cells and the demyelinating processes that cause the MS symptoms.

II. MS symptoms

Besides the pathological markers, currently MS is diagnosed and classified based on the spontaneous presentation of symptoms that appear suddenly over the course of several days, last for usually short periods of time (from a few days to a few months), until complete or at least partial recovery, and remain symptom-free for months or years. These spontaneous attacks are defined as relapses and the recoveries as remissions (Miller, 2001). Each new relapse results from an attack on the white matter of the central nervous system, and depending on its location in the areas of the brain or the spinal cord, it can cause very different symptoms. During a relapse the patient may experience new symptoms or an increase in existing symptoms.

The most common MS physical symptoms experienced by the patients consist of weakness; stiffness; paralysis; tremor; fatigue; disturbances in coordination, gait and vision; difficulties swallowing and speaking; disruption of bladder and bowel functions, and sexual dysfunction; sensory changes and heat sensitivity; and psychiatric and cognitive problems (Miller, 2001). The expression of MS symptoms is very heterogeneous, causing the patients to suffer from a very individual mixture of various symptoms that range in their order of accumulation and severity as well.

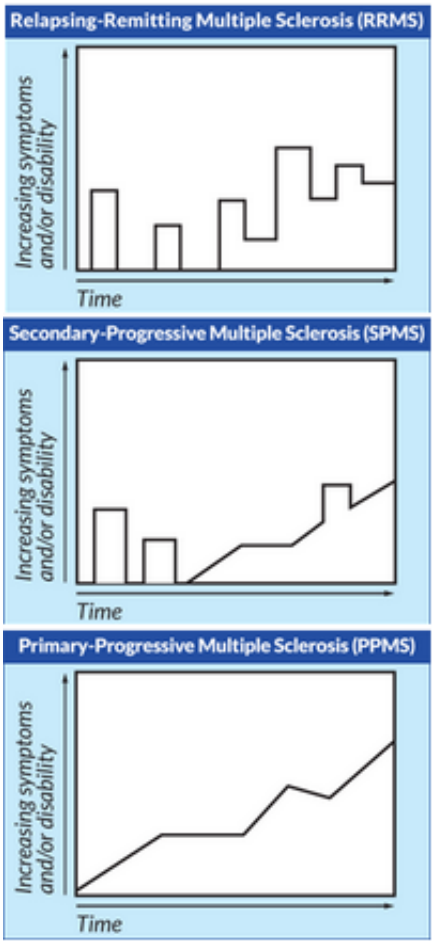
1.1.3. MS subtypes

Due to the heterogeneity and individual variation of MS symptoms, in clinical practice the patients are grouped based on their MS subtype rather than on their symptoms (Chelune, Stott, & Pinkston, 2008). The classification of MS subtypes is based on consensus and relies on the clinical descriptors of the course of the disease based on current status and history. Nowadays MS is most often classified into relapsing-remitting (RRMS), secondary-progressive (SPMS), and primary progressive (PPMS)

(Lublin et al., 2014). The disease onset and clinical course are highly variable and mostly unpredictable.

The majority of patients (around 80%) start with a RRMS course, which is characterised by a pattern of clinical attacks (relapses) from which the patient essentially recovers (remissions), and there's relative stability between the attacks (Figure 1.1). Few patients with RRMS do extremely well, with seldom exacerbations and little deterioration, but the majority (around 60%) of people with RRMS after a while start accumulating symptoms progressively, either with or without occasional relapses. This subtype of MS is termed SPMS.

Figure 1.1. An illustration of relapses and progression of MS symptoms in RRMS, SPMS and PPMS over time



Note. Image adapted from Multiple Sclerosis Association of America website (http://mymsaa.org/about-ms/overview/#TypesofMS ; retrieved on 21st July, 2015)

However, there's a subgroup of patients (around 15%) who develop PPMS. This course of MS is defined by continuous worsening of symptoms and progressing disability without distinct relapses. The beginning of PPMS is typically around age 40, which is around 10 years later than the start of RRMS, but around the same age as the transition from RRMS to SPMS happens (Thompson et al., 2000). On occasions, PPMS progresses very rapidly, reducing a patient to helpless dependency or death soon after disease onset.

Besides the relatively common RRMS, SPMS and PPMS, other type of MS exists, called progressive-relapsing MS (PRMS). PRMS is characterized by steadily progressing disease from the beginning and occasional exacerbations along the way. People with this form of MS may or may not experience some recovery following these attacks; the disease continues to progress without remissions (Lublin, et al., 2014).

Another term commonly used in MS staging is the clinically isolated syndrome (CIS) (Lublin et al., 2014). CIS is recognised as the first clinical presentation of the disease that shows characteristics of inflammatory demyelination that could be MS, but has yet to fulfil criteria of dissemination in time. However, PRMS and CIS are much less common and most research focuses on RRMS, SPMS and PPMS types only.

1.2. Cognitive dysfunction in MS

In addition to the physical symptoms, around 45- to 65-percent of pwMS experience cognitive symptoms as well. These typically involve deficits of speed of information processing; attention and concentration; memory; and executive functions (DeSousa, Albert, & Kalman, 2002). Although cognitive dysfunction is one of the most prominent symptoms associated with MS as it occurs in over half of the patients, it is often less severe than those seen in other neurodegenerative disorders in which dementia is prominent (such as Alzheimer's disease (AD) or Fronto-temporal dementia (FTD) (Filley, Heaton, Nelson, Burks, & Franklin, 1989). In MS cognitive dysfunction is probably the most important determinant of employment status and associated societal costs, and also adversely affects everyday tasks, such as driving safety; household task completion; social activity; physical independence; rehabilitation progress; coping; treatment adherence; and mental health

(Langdon, 2011). Furthermore, patients' cognitive impairment is one of the leading sources of caregiver strain (Chipchase, 2001), making other MS symptoms more difficult to manage.

Data on the prevalence of cognitive deficits in pwMS varies with MS subtypes and tests performed (Chelune, Stott, & Pinkston, 2008), and the basic constituents of cognitive dysfunction in multiple sclerosis are still under debate. The frequency and pattern of cognitive functions affected vary considerably between and within patients, since it is hypothesized to typically consist of differently intermingled domain-specific deficits rather than of a uniform overall cognitive decline. This could potentially be due primarily to the heterogeneity of the disease in its extent, its location and the dynamics of its pathological processes (Hoffmann, Tittgemeyer, & von Cramon, 2007).

Because of the many ways in which the lesions can manifest in the brain, pwMS can have their cognition affected in many ways, therefore this is why a neuropsychological examination of these patients requires assessment of a variety of functions (Lezak, 2004). Due to the heterogeneity and lack of clear definition of cognitive deficits in MS, patients seen in MS clinics and neurologic practices are not routinely assessed neuropsychologically.

1.2.1. Information processing speed and attention

Many pwMS report feeling mentally slowed down, and find it difficult to concentrate and keep up with the pace of normal conversations. Measures of information-processing speed appear to be the most robust and sensitive markers of cognitive impairment in multiple sclerosis patients. Recent studies demonstrate that single, predominantly speed-related cognitive tests may be superior to extensive and time-consuming test batteries in screening overall cognitive decline (Hoffmann et al., 2007).

Impaired processing speed is a classic finding in MS, irrespective to the modality of stimulus presentation (auditory or visual) (DeLuca, Chelune, Tulsky, Lengenfelder, & Chiaravalloti, 2004), and is considered to be independent of motor involvement (Rao, Aubin-Faubert, & Leo, 1989). Moreover, other studies have found that information processing speed deficits tend to be separate from attentional deficits as MS patients can often perform at the same level of accuracy as the

controls if the stimuli are presented at a sufficiently slow rate or if the task is self-paced (Demaree, DeLuca, Gaudino, & Diamond, 1999; Lengenfelder et al., 2006).

Processing speed is often considered to be the initial symptom of MS-related cognitive deficit, that later with progression of the disease contributes to working memory and long-term memory impairment in pwMS (Archibald & Fisk, 2000; Litvan, Grafman, Vendrell, & Martinez, 1988). Some MS-type differences exist with general slowing appearing to be highest in the progressive types of MS (50%) and less prominent in the RRMS (24%) (De Sonneville et al., 2002).

Regardless of disease status, most MS patients exhibit deficits on attention tasks with greater stimulus or response complexity, and those requiring inhibition of previously correct response (Paul, Beatty, Schneider, Blanco, & Hames, 1998), shifting attention back and forth from one stimulus to another (Grigsby, Kaye, & Busenbark, 1994) and dual-tasking (D'Esposito et al., 1996).

1.2.2. Memory and learning

PwMS often report problems with explicit memory, such as remembering recent events, times of appointments, and names of people and objects. This deficit is more apparent for recent events rather than events from distant past and their semantic memory is often well preserved (Prakash, Snook, Lewis, Motl, & Kramer, 2008). Moreover, free recall is reported to tend to be poorer than cued recall, which in turn to be inferior to recognition (Thornton & Raz, 1997). Therefore the memory impairment among pwMS is argued to be a consequence of inadequate initial learning and not a function of impaired retrieval (Chiaravalloti, Balzano, Moore, & DeLuca, 2009; DeLuca, Barbieri-Berger, & Johnson, 1994). As previous studies suggest, the impaired encoding capacity could be partly explained by poor ability to activate mnemonic strategies, such as semantic clustering (Arnett et al., 1997) or visual imagery (Canellopoulou & Richardson, 1998).

1.2.3. Executive functions

MS patients are repeatedly reported to have problems with tasks involving cognitive estimation, planning, sequencing and problem solving (Arnett, et al., 1997; Foong et al., 1997). These deficits detrimentally affect many aspects of daily life, such as the ability to run a household, participate in

social events, and maintain employment—factors that can all affect the overall quality of life of the patient. Disorders of executive functioning commonly are linked to behavioural problems, such as apathy or disinhibition, and are often more apparent to family members than they are to the affected individual, resulting in persons close to the patient erroneously attributing these behaviours to personality features, such as disorganization or stubbornness (Benedict et al., 2000; Chiaravalloti & DeLuca, 2003), causing even more issues in social interactions at home or at work.

Verbal fluency is often disrupted in MS, whether by reductions in processing speed, flexibility, search strategy, or access to semantic storage, and suggesting that measures of verbal fluency may be amongst the most sensitive neuropsychological measures to cognitive impairment in MS (Henry & Beatty, 2006). In addition, some studies report that phonemic fluency tasks are more sensitive to impairment than semantic category fluency tasks (Connick, Kolappan, Bak, & Chandran, 2012), however, in MS samples semantic category fluency has been more widely studied.

1.2.4. Visuoperceptual functions

Eye problems such as blindness, partial vision, blurring, or decrease in visual acuity, visual fields and colour perception, are common in MS and they are often caused by damage to the optic nerve. Besides problems with the peripheral vision, pwMS tend to have complaints about their vision that relate to damage to the higher level processing of visuoperceptual information. Any aspect of visuoperception may be disrupted, including facial perception ("knowing who"), visual form perception ("knowing what") and visuospatial perception ("knowing where") (Rao, Leo, Bernardin, & Unverzagt, 1991; Vleugels et al., 2000).

For example, pwMS often report problems with processing visual information from reading and watching the television where they report being able to see the images, but to have difficulties processing their meaning. A more dangerous example is from family members expressing concern about patients who have visuospatial difficulties and tend to drive too close to the side of the road and have some "near misses" on the highway (Fischer, 2001).

1.2.5. Language functions

In MS language functions are considered to typically remain intact except for those dependent on rapid and efficient retrieval (Lezak, 2004). Deficits in confrontation naming have been reported in MS, however, those deficits could be argued to be caused by disruption in retrieval rather than semantic storage, as phonemic cuing often facilitates retrieval in MS (Prakash, et al., 2008).

Dysarthrias (difficulties articulating) are rare in MS, and if such severe language disturbances occur, they tend to be associated with damage to the brainstem and inability to perform movements involving the mouth, such as swallowing (Kurtzke, 1983), rather than linguistic problems. However, subtle language problems do tend to occur in MS, such as difficulties with comprehension of complex or ambiguous grammatical structures, which causes problems with keeping up with conversations (Grossman et al., 1995). Moreover, pwMS have been reported to have the tendency to generate verbal output with fewer information units per sentence, and fewer complete and grammatically correct sentences (Wallace & Holmes, 1993). These small difficulties in sentence generation can cause frustration and have detrimental effects on inter-personal communication at home and on work performance where the job involves verbal communication.

1.2.6. Prevalence of impairment on different cognitive domains

Currently it is estimated that around 40% of pwMS in the community samples (Rao et al., 1991) and around 50-60% of pwMS in hospital samples (Ron, Callanan, & Warrington, 1991) have some level of cognitive impairment. An estimate of prevalence of cognitive deficits can be seen in Table 1.1.

Table 1.1. Prevalence of MS-related impairment on cognitive functions

Domain of cognitive function	Prevalence of severe impairment [a]
Most commonly impaired	
Episodic memory	22-31%
Complex attention/ processing speed	22-25%
Verbal fluency	22%
Impaired moderately often	
Executive functions	13-19%
Visual perception	12-19%
Less frequently impaired	
Language/ semantic memory	8-10%
Attention span	7-8%

[a] Severe impairment – performance below the 5th centile of demographically matched healthy controls. Table taken from Fisher (2001), with its contents adapted from Rao et al. (1991)

1.2.7. Consensus approach to domain-specificity of cognitive impairment in MS

Since the reported deficits in cognitive functioning of pwMS are highly heterogeneous, there have been systematic approaches to generalize and simplify the concept of MS-related cognitive impairment.

The first attempt to systematically review current studies into cognitive deficits in MS was published by Janis M. Peyser, Stephen M. Rao, Nicholas G. LaRocca and Edith Kaplan in 1990. In their seminal paper they identified a problem that part of the reason why it was so difficult to define MS-related cognitive impairment, was the heterogeneity of research methods and samples employed in the neuropsychological studies. Therefore they aimed to uniform the methods in the research area by proposing a battery comprising of seventeen tests to assess all of the cognitive functions considered to be related to MS (Peyser, Rao, LaRocca, & Kaplan, 1990), including information processing speed, attention/concentration, memory, language, visuoperceptual and executive functions. In his attempts to shorten this battery, Stephen Rao has chosen to leave only the tests which he found most sensitive to MS, resulting in a five item test battery measuring verbal and visuospatial learning and memory, information processing speed, sustained attention, and verbal fluency (Rao et al., 1991). He named

his battery Brief - Repeatable Battery of Neuropsychological tests (BRBN) (Rao, 1990) and since then this battery is considered to be the gold standard of cognitive assessment in MS.

Due to the popularity of the BRBN and its subtests, the cognitive domains it measures have been well studied in MS. Much less is known about the cognitive domains not covered by the BRBN and its variants, such as language and visuoperceptual functions.

The BRBN was comprised to be a compilation of the tests proven to be most sensitive to MS-related cognitive impairment, and to save time three of its components have been used as standalone assessments in many research studies. The Paced Auditory Serial Addition Test (PASAT), Symbol-Digit Modalities Test (SDMT) and the verbal fluency tasks measure predominantly speed of information processing, but at the same time address other complex functions such as memory, attention, and executive functions. It is important to note that these most sensitive and the most often used tests are, in a way, 'dirty tests' as they assess a compound of various cognitive functions. In a clinical setting, such 'dirty tests' serve physicians as a valuable tool to differentiate patients with and without cognitive impairment, but not as much to identify which functions are impaired in individual patients in order to define starting points for individualized disease management (Hoffmann et al., 2007). Since MS-related cognitive impairments are highly heterogeneous, it could be considered insufficient to administer solely the PASAT, SDMT and category fluency tasks in studies addressing cognitive deterioration.

1.3. Links between cognitive dysfunction and disease variables

Depending on a study, cognitive dysfunction is considered to be partly explained by MS type, MS duration, neurological disability, brain tissue damage, levels of depression, and cognitive reserve. However, depending on their methodology, many studies tend to present discrepant findings.

1.3.1. MS type

Even though the classification into RRMS, SPMS and PPMS is primarily based on progression of physical dysfunction, some courses of MS are more associated with cognitive difficulties than others. In general, patients with RRMS tend to outperform patients with PPMS and SPMS on cognitive

examinations (Grossman et al., 1994), and PPMS patients tend to perform better than those with SPMS (Camp et al., 1999; Gaudino, Chiaravalloti, DeLuca, & Diamond, 2001). RRMS patients may also have cognitive deficits when compared to healthy controls, albeit less obvious ones than those observed in SPMS and PPMS (Grossman et al., 1994; Ryan, Clark, Klonoff, Li, & Paty, 1996).

However, keeping that in mind, the relationship between cognitive impairment and disease course is not strong enough to predict the cognitive status of individual MS patients (Beatty, Goodkin, Hertsgaard, & Monson, 1990), especially once the patients are equated for disease duration and disability (Foong et al., 2000).

1.3.2. Disease duration

Among the factors playing a role in development of cognitive dysfunction in MS, disease duration is reported not to play a significant role (Beatty, et al., 1990; Lynch, Parmenter, & Denney, 2005). In the course of a sufficiently long follow-up cognitive dysfunction is likely to emerge and progress in a sizable proportion of patients, however, it is likely to reflect the progression of disease and development of more neurologic symptoms and brain tissue damage. The number of years someone has had MS alone is not a sufficient predictor of cognitive symptom development (Amato, Ponziani, Siracusa, & Sorbi, 2001).

1.3.3. Neurological disability

The neurological disability as indicated by the score on the Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) alone fails to account for the progression of cognitive deficits (Beatty, et al., 1990); however correlations are reported between measures of neurological disability and tests of information processing speed (Lynch et al., 2005).

Studies show relapses to be associated with fluctuations in cognitive function, as during a relapse the patients tend to perform more poorly on cognitive tests, particularly on those involving attention and processing speed (Foong et al., 1998); and the total number of relapses since diagnosis is shown to correlate with overall cognitive dysfunction (MacAllister et al., 2005).

1.3.4. Brain tissue damage

Studies involving factors that contribute to the severity of cognitive impairment indicate, that the extent of tissue damage in white and grey matter tend to be among the best correlates. Brain atrophy has been reported to account for more variance than lesion burden in predicting cognitive impairment in MS (Filippi et al., 2010), and central atrophy in the thalamus in particular is strongly associated with neuropsychological morbidity, as it is believed to mediate cognitive function via cortical and subcortical pathways (Benedict et al., 2004).

A considerable number of imaging studies have attempted to identify the relationship between the location of multiple sclerosis lesions and modality of cognitive dysfunction, but no conclusion was reached as the studies produced conflicting results, which can partially be explained by the heterogeneous pathological substrate of multiple sclerosis lesions and the fact that, besides the periventricular white matter, there is no preferred region where lesions aggregate (Hoffmann et al., 2007).

1.3.5. Depression

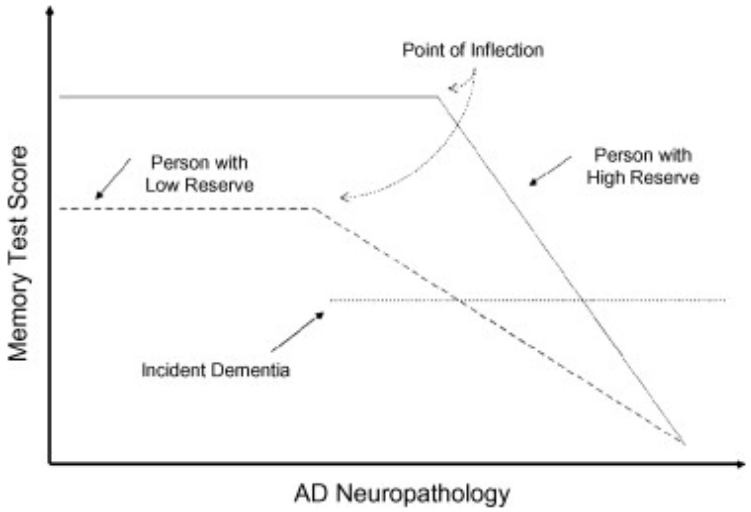
Depressed mood in MS patients has been reported to be associated with poorer performance on processing speed, executive (Arnett, Higginson, & Randolph, 2001) and working memory tests (Arnett, Higginson, Voss, Bender et al., 1999). The authors argue, that the impaired performance on speeded attentional tasks of the depressed mood pwMS could be potentially caused by several factors, such as reduced working memory capacity, impaired deployment of executive strategies, or psychomotor slowing (Arnett et al., 1999). Furthermore, there is evidence that depression causes brain pathology (i.e. causes changes in hippocampal structure) (Videbech & Ravnkilde, 2004), which makes in turn it difficult to dissociate from MS-related pathology.

However, it has been shown that the detrimental effects of depression on cognition could in part be reversible with exercise in interventional studies (Motl & Pilutti, 2012; Stroud & Minahan, 2009; Velikonja, Čurić, Ožura, & Jazbec, 2010), therefore it could be considered that the effects of depression may be mediated by individual differences and lifestyle variables.

1.3.6. Cognitive reserve

Cognitive reserve is a relatively new and understudied concept in MS, however, successful models of cognitive reserve have been developed and studied in other neurodegenerative conditions, namely the Alzheimer's Disease. The idea of reserve stems from the observed lack of direct relationship between the degree of brain pathology and its clinical manifestation (Stern, 2002). For example, the same magnitude of brain damage caused by neurodegeneration can result in varying levels of cognitive impairment in different individuals, depending on their level of reserve (Stern, 2009; Figure 1.2). Specifically, the point of inflection, where the cognitive functioning begins to be affected by pathology, starts later in individuals with higher levels of reserve, and in result the clinical manifestation of cerebral pathology is reached later, when pathology is more severe.

Figure 1.2. Theoretical illustration of how reserve may mediate between brain pathology and Alzheimer's Disease clinical expression



Note. This figure has been adapted from Stern, 2009.

Literature into neurodegenerative disorders distinguishes between two types of reserve: passive and active. Brain reserve is an example of a passive model, where the protective resources derive from the brain size or neuronal count. Specifically, it has been shown that larger brains can sustain more

atrophy before the clinical deficits emerge, because sufficient neural substrate remains to support normal function (Katzman, 1993). On the other hand, active models, such as cognitive reserve, suggest that the brain actively attempts to cope with brain damage by using pre-existing cognitive processes or by enlisting compensatory processes (Stern, 2002). Therefore according to the reserve theory, although two patients may have the same amount of passive (brain) reserve capacity, the patient with more active (cognitive) reserve would better sustain brain damage and maintain function for longer.

Previous work into cognitive reserve indicates that a set of life experiences, such as educational and occupational exposure and leisure activities, are associated with slower rate of cognitive decline and lower rates of developing dementia; and that each of these life experiences tend contribute to cognitive reserve independently (Stern et al., 1994). Of the methods which could be used in research as proxies for cognitive reserve, educational exposure is the most commonly employed; however, even though less often studied, engagement in leisure activities has been shown to be a better indicator of protective factors associated with cognitive reserve (Valenzuela & Sachdev, 2005).

Most of the work on cognitive reserve has been done on cognitive ageing and people with dementia, however, life experience proxies for higher cognitive reserve have also been reported to mediate cognitive changes associated with HIV dementia (Farinpour et al., 2003), schizophrenia, bipolar disorder and depression (Barnett, Salmond, Jones, & Sahakian, 2006), and traumatic brain injury (Kesler, Adams, Blasey, & Bigler, 2003).

Previous work involving MS patients has shown that pwMS with higher reserve tend to be better able to withstand MS neuropathology without cognitive impairment (Sumowski, Chiaravalloti, Wylie, & DeLuca, 2009). Specifically, in pwMS higher premorbid ability and more years of education have been reported to account for slower deterioration of information processing speed (Benedict, Morrow, Weinstock Guttman, Cookfair, & Schretlen, 2010) and cerebral tissue destruction (Bonnet et al., 2006), which potentially is indicative of cognitive compensation in more educated patients. Cognitive reserve or lifestyle choices (indicated as early-life cognitive leisure) were shown to be protective against MS-related cognitive impairment independently of genetic factors outside of one's control, and the effect of cognitive reserve was reported to be superior to

brain reserve (indicated as larger maximal lifetime brain volume) (Sumowski et al., 2013). Interestingly, cognitive reserve related compensation is reported to be most beneficial at earliest stages of MS and may, however, fail with progression of damage (Amato et al., 2013).

1.3.7. Exercise

Among health behaviours that have been shown to have a compensatory effect on progression of cognitive deficits, exercise is reported to exert a prophylactic effect for MS progression. Previous research from observational studies have reported a positive association between cardiorespiratory fitness and structural brain volumes and preservation of neuronal integrity (Prakash, Snook, Motl, & Kramer, 2010). In addition, observational studies have shown that pwMS who regularly exercise tend to maintain better affective overall mental health (Turner, Kivlahan, & Haselkorn, 2009), have less long-term disability and perform better on measures of processing speed (Prakash, et al., 2010). However, the results from interventional studies limit definite conclusions from the observational studies on the disease-modifying effect exercise has on clinical measures of MS, but the effect on MRI, animal model and patient self-reported feeling of well-being has been well supported (Dalgas & Stenager, 2012).

1.4. Self-awareness of cognitive deficits in MS

Much research has been done in order to identify and explain the course of MS progression and new cognitive deficit acquirement. However, less is known about the patients' subjective perception of their deficit progression. Based on the previous literature into MS and into other neurodegenerative illnesses there are grounds to speculate that pwMS might have difficulties in their self-perception of cognitive deficit progression.

Disordered awareness of cognitive and behavioural deficits, or anosognosia, is one of the leading factors determining the prognosis, the development of coping strategies and the efficacy of post-diagnostic support (Robinson et al., 2011). Underestimation of existing deficits can have a negative impact on patients' engagement with post-diagnostic support, and can lead to performing dangerous or difficult activities beyond one's capabilities (Yokoi & Okamura, 2013). This could include taking part in such activities as cooking, driving or getting around independently, potentially leading to

accidents. If identified early, to some degree the awareness deficits can be managed, possibly avoiding some of the potential mishaps.

Evidence from neuroimaging studies suggest an association between impairments in aspects of self-awareness and insight and frontal and prefrontal lobe neural atrophy in neurodegenerative disorders (Rosen et al., 2011; Stuss et al., 2000). Loss of insight is considered to be one of the core features of frontal/behavioural variant of Fronto-Temporal Dementia (FTD) (Eslinger et al., 2005), however, due to many shared clinical and pathological features, it has also been reported and studied, although to a lesser extent, in other neurodegenerative conditions, such as Alzheimer's Disease (AD) (Robertsson, Nordström, & Wijk, 2007), Parkinson's Disease with dementia (PD) (Prigatano, Maier, & Burns, 2010), Progressive Supranuclear Palsy (PSP) and Corticobasal Degeneration (CBD) (O'Keefe et al., 2007). From studying insight in FTD, PSP, CBD and healthy controls, O'Keefe et al. (2007) concluded that awareness problems are caused by involvement of the frontal and prefrontal regions, and could be predicted by poorer performance on frontal-executive tests, such as set-shifting, planning and categorisation (Grafman & Litvan, 1999).

Since pwMS tend to have more extensive cortical atrophy in superior frontal and parietal cortices than in other regions (Benedict et al., 2002), and, executive functions are reported to be impaired moderately often in pwMS (13-19%; Table 1.1.) (Rao et al., 1991), under the same reasoning it could be hypothesised that pwMS may be vulnerable to experiencing deficits in insight as well.

And indeed, even though neuropsychological evaluations have estimated some degree of cognitive impairment to be prevalent in around half of pwMS, however, the outcomes of specialist assessments don't always go hand in hand with patient self-reports of their cognitive deficits. In one study pwMS have been found to not be fully aware of their current cognitive deficits (or not report them reliably) when examined cross-sectionally, however, when examined longitudinally, pwMS were found to be accurate in their reports of perceived levels of cognitive change (Christodoulou et al., 2005).

Another study that investigated the self-perception of cognitive deficits in MS has shown that the pwMS could be classified according to whether they overestimated or underestimated their cognitive ability. However, instead of comparing the patients' self-reports to their actual performance, this

study compared them to informant ratings (usually spouse or family member). In this study, compared to underestimations, overestimations of performance were linked to lower levels of depression and conscientiousness, and greater degrees of cognitive impairment, euphoric behavioural disinhibition, and unemployment (Carone, Benedict, Fishman, & Weinstock-Guttman, 2005). This again highlights the behavioural consequences of insight deficits in pwMS.

The exploration of awareness deficits is a relatively new and unexplored field of study in MS. Most research of awareness deficits in clinical samples have been performed on patients with dementia, brain injury, and stroke. Research into those samples has shown that metacognition can be affected differently for each cognitive domain (i.e. for example, the patient may be aware of disruption in memory, but not aware of disruption of executive functions), therefore the awareness of cognitive deficits in each cognitive domain should be assessed separately (Cosentino & Stern, 2005; Schoo, van Zandvoort, Biessels, Kappelle, & Postma, 2013).

1.5. Prognosis and cognitive symptom progression in MS. Longitudinal studies

Great variability in longitudinal changes in cognitive performance in pwMS exists with some patients performing at the same level with controls even after many years with MS, and other patients' cognitive symptoms progressing rapidly from the very start of their MS (Benedict & Zivadinov, 2011; Staff, Lucchinetti, & Keegan, 2009). Predicting the probability and course of cognitive impairment in an MS patient is difficult, as it may depend on many different variables, such as genetics, male sex, progressive disease course, MRI metrics and comorbid neuropsychiatric illnesses (Benedict & Zivadinov, 2011; de Groot et al., 2009). Furthermore, slowed progression of neuropsychological deficits is also reported to be linked to health behaviours and cognitive reserve (i.e. higher level of premorbid cognitive functioning) (Benedict et al., 2010; Sumowski et al., 2009).

Since around half of people with MS suffer from some degree of cognitive problems, there's a number of longitudinal cohort studies that investigate how the cognitive symptoms develop and progress in a sample of pwMS. Most longitudinal studies in MS were developed to have a retrospective design, where the neuropsychological evaluation is performed only at follow-up with only the predictor variables, such as MRI or clinical course, collected at baseline. In order to

investigate the heterogeneous progression of cognitive deficits in pwMS, the cohort approach where the same individuals undergo neuropsychological assessments at baseline and at follow-up, separated by a time period, can be deemed superior to retrospective research. Therefore for the purposes of this study it was chosen to review only the adult cohort longitudinal studies where the cognitive evaluation was performed at least twice by a specialist researcher, and the baseline and follow-up phases are clearly defined and separated by a time interval. Studies using survey design or self-reported measures of cognitive dysfunction, or longitudinal studies designed to validate cognitive assessment tools or interventions, rather than investigate naturally occurring cognitive changes, were not deemed suitable for the purposes of this project and therefore were not reviewed.

To the best of my knowledge 23 existing longitudinal cohort studies investigating cognitive changes in pwMS can be identified (Table 1.2.). However, the results from these studies are difficult to compare, since these studies differ substantially in their methodologies; clinical characteristics of the patients; follow-up intervals; cognitive domains assessed and the neuropsychological measurements employed per domain; cut-off points in the definition of cognitive impairment; follow-up rates; and methods of statistical analyses (Table 1.2.). In addition, the majority of longitudinal studies do not employ a control sample (or if they do they only assess them once as a cut-off reference point without following them up longitudinally), therefore they do not account for the longitudinal changes in cognitive abilities that are related to normal variability in performance, cognitive ageing, practice effect, or systematic differences in administration, rather than being caused by MS. High drop-out rates in some studies are also problematic as they indicate poor representation of the sample and the population and in turn make the findings difficult to interpret and generalize.

In addition to differing methodologies, some studies present discrepant findings, especially on the modest effect that the MS type, disease duration and neurological disability have on cognitive deterioration. Strongest correlations were found in the studies where pwMS were classified into ‘cognitively impaired’ and ‘cognitively unimpaired’ at baseline. The ‘cognitively impaired’ tend to have higher WML load and more brain atrophy and poor cognitive performance at baseline, and their cognitive symptoms progress more rapidly, irrespective of the MS type. These people can be identified early in their MS by poorer performance on information processing speed tasks, which

later leads to moderately worsening performance in memory and executive domains, and this process occurs at a faster pace than for the 'cognitively unimpaired' group, and is considered to be independent of progression of neurological disability.

The reason why some pwMS fall into 'cognitively impaired' and 'cognitively unimpaired' groups at baseline is unclear, and could potentially have to do with general intellectual ability or cognitive reserve at baseline, and the rate of cognitive symptom progression could be mediated by certain health behaviours.

Currently in the MS literature the research on the progression of deficits measured by the BRBN battery (verbal and visuospatial memory, processing speed, attention/concentration, and verbal fluency) is overrepresented, thus lacking a systematic approach to cognitive deterioration in temporal and posterior regions, such as language and visuoperceptual abilities. All (100%) of previous studies included at least one measurement of memory, 87% included measures of attention, 65% included information processing speed or executive functioning, and only 30% of studies have looked into visuoperceptual functions or language (Table 1.2.). I found only one study that looked into all cognitive functions (Strober et al., 2014), but this study had an extremely low follow-up rate (29.34%), therefore their findings can be considered unfit for generalizations.

The heterogeneity of methodologies employed and cognitive tests used make the dimensionality of cognitive impairment in MS difficult to establish. Some studies treat cognitive impairments as one entity and do not dissociate between the mechanisms that cause deterioration in separate functions, while other studies claim that MS causes disruptions in information processing speed, which later results in development of problems in other cognitive domains.

Table 1.2. Summary of prospective non-interventional longitudinal studies of cognition in MS arranged by total follow-up interval

Study	Sample size with full data	Follow-up interval	Cognitive functions assessed						Follow-up rate	Predictors/correlates of cognitive decline
			I	A	M	E	V	L		
(Morrow et al., 2009)	51 RRMS, 17 SPMS, 1 PPMS, 70 controls	1.5 year	✓	✓	✓	✓	✓		25%	Cognitive impairment starts with visuo-perceptual deficits, and then with disease progression spreads into memory, executive deficits and processing speed
(Mariani et al., 1991)	19 RRMS	2 years			✓	✓	✓		N / A	No longitudinal cognitive decline observed; only individual differences
(Zivadinov et al., 2001)	53 recently diagnosed RRMS	2 years		✓	✓	✓		✓	N / A	Early axonal loss and brain atrophy, but not clinical variables, indicate the start of cognitive degeneration
(Camp et al., 2005)	99 PPMS	2 years	✓	✓	✓	✓			67.3%	Lower baseline cognitive status and later accumulation of WML load
(Duque et al., 2008)	6 CIS, 22 RRMS, 7 SPMS, 4 PPMS, 25 controls	Every 3 months for 2 years	✓	✓	✓				89%	Information processing speed deficits correlate with disease progression and disability, but memory problems occur after 2 years irrespective of disease progress
(Simioni et al., 2009)	70 RRMS	2.1 years		✓	✓	✓			64%	Decline in decision making appears as an isolated deficit early in MS and is not related to clinical variables
(Kujala, Portin, & Ruutiainen, 1997)	42 pwMS, 34 controls	2.8 years	✓	✓	✓	✓			88%	Phenotyping into 'cognitively preserved' and 'cognitively deteriorated' at baseline
(Patti et al., 1998)	34 SPMS, 23 RRMS	3 years		✓	✓	✓			N / A	WML load and physical disability moderately correlate with cognitive deterioration
(Arnett, 2005)	31 RRMS, 15 SPMS, 6 PPMS, 1 PRMS	3 years	✓	✓	✓	✓			69%	Self-reported depression predicts cognitive dysfunction
(Denney, Lynch, & Parmenter, 2008)	24 PPMS, 25 controls	3 years	✓		✓	✓			67%	Slowed information processing speed

Table 1.2. (continued)

Study	Sample size with full data	Follow-up interval	Cognitive functions assessed						Follow-up rate	Predictors/correlates of cognitive decline
			I	A	M	E	V	L		
(de Groot, et al., 2009)	146 pwMS	0.5, 1, 2 and 3 years	✓	✓	✓				93.6%	Older age, male sex, perceived ability to concentrate and WML
(Jennekens-Schinkel et al., 1990)	13 RRMS, 20 PMS, 18 controls	4 years	✓		✓		✓	✓	85%	Clinical variables cannot explain cognitive deterioration
(Sperling et al., 2001)	28 pwMS, 28 controls	1 and 4 years	✓	✓	✓				64%	WML volume in frontal and parietal regions
(Reuter et al., 2010)	24 CIS, 13 controls	5 years	✓	✓	✓	✓		✓	N / A	WML load at baseline
(Glanz et al., 2012)	90 pwMS	1, 2, 3, 4 and 5 years	✓	✓	✓				70%	Slowed visual information processing speed
(Altinkaya et al., 2012)	20 RRMS, 5 SPMS	5.3 years		✓	✓	✓	✓	✓	55.6%	Disease duration for some functions
(Haase et al., 2004)	20 RRMS, 27 controls	7 years		✓	✓	✓	✓	✓	74%	Cognitive dysfunction correlates to disease duration but not disability
(Deloir et al., 2010)	45 RRMS, 56 controls	1, 2, 5 and 7 years	✓	✓	✓				78%	Phenotyping into 'cognitively unimpaired' and 'cognitively impaired', and diffused brain damage at baseline; and progressive central brain atrophy over 2 years after diagnosis
(Bergendal, Fredrikson, & Almkvist, 2007)	10 RRMS, 17 SPMS, 4 PPMS	8 years	✓	✓			✓		97 %	Slowed processing speed and high disability in SPMS patients at follow-up
(Piras et al., 2003)	12 RRMS	8.5 years		✓	✓	✓			37.5%	Slowed processing speed, WML and cortical atrophy
(Schwid et al., 2007)	153 pwMS	2 and 10 years	✓	✓	✓				69%	Rate of deterioration in first two years was predictive of 10 year change
(Amato et al., 2001)	44 RRMS, 6 PPMS, 70 controls	4 and 10 years		✓	✓	✓		✓	90%	Neurological disability, progressive MS course and older age
(Strober et al., 2014)	22 pwMS	17.9 years	✓	✓	✓	✓	✓	✓	29.34%	Drop in processing speed
			I	A	M	E	V	L		
% of longitudinal studies with at least one test per that cognitive domain			65	87	100	65	30	30		

Abbreviations: RRMS - Relapsing-Remitting MS, SPMS - Secondary Progressive MS, PPMS - Primary Progressive MS, PMS - progressive MS, PRMS - Progressive Relapsing MS, CIS - Clinically Isolated Syndrome. I - Information processing speed, A - Attention, M – Memory, E - Executive, V – Visuoperceptual, L - Language functions. WML – White matter lesions

To the best of my knowledge no study up to date has analysed the dimensionality of cognitive impairment to investigate the pattern of how MS-related cognitive decline starts and spreads, while accounting for inter-personal variation that pwMS exhibit: MS course and levels of disability, baseline intellectual ability, cognitive reserve and exercise, levels of depression and various medication. In addition, no study so far has systematically examined the patients' perception of cognitive decline in MS samples, and related their self-perception of change to the actual longitudinal changes in performance on neuropsychological tests. I hope that with such study the discrepancies in findings from previous research could be merged and explained and if a sufficiently large and heterogeneous sample is collected. Such investigation could also broaden the current understanding of MS-related progression of cognitive deficits whilst accounting for individual variation, which normally causes problems when trying to compare findings from different longitudinal studies.

1.6. Gaps of knowledge and predictions

1.6.1. Development of the scientific framework for the longitudinal study

This project aimed to set up the scientific framework through which a sample of pwMS could be followed in order to study the MS-related longitudinal change between the baseline and follow-up assessments. The scientific framework for this project was developed based on the previously published literature that addresses longitudinal cognitive assessment in MS. During this process several gaps of knowledge were identified, and predictions were made about the development and progression of cognitive impairment in pwMS, its dimensionality, and clinical and non-clinical correlates. This project was devised to fill these identified gaps of knowledge and to test the following informed predictions.

1.6.2. Predictions for dimensionality of cognitive impairment in MS

The first identified gap of knowledge revolved around the dimensionality of MS related cognitive impairment. Based on the previous literature separate hypotheses were made for cross-sectional and longitudinal analyses.

I. Cross-sectional predictions

The first goal of this project was to investigate whether MS affects all of the cognitive functions or only some of them. It was predicted that due to heterogeneity of brain lesion locations, MS has the potential to affect all cognitive domains, although independently and not necessarily at the same frequency. Based on the small body of existing literature the cognitive deficits are predicted to occur in recognizable patterns, and potentially groups of pwMS who have shared patterns of cognitive test failure can be identified.

II. Longitudinal predictions

The next step was to identify the longitudinal change in the dimensionality of cognitive deficits in pwMS. If we accept the premise that pwMS have impaired cognition because they have pathology affecting their brains, and that the extent (burden) of this randomly distributed pathology will increase over time, then we can predict that the pattern of cognitive deficits will become more homogeneous (and unidimensional / 'global') over time at the population level, reflecting increasing multidimensionality at the individual level.

1.6.3. Predictors of longitudinal change in cognition

The second aim of this project was to investigate what causes longitudinal change in cognitive functioning.

A number of predictions can potentially be made about the effects of clinical and non-clinical variables. Based on the review of the previous literature, the following model can be regarded as plausible:

$$\textit{Longitudinal change in cognition} = \alpha + \beta_1\,(\textit{clinical factors}) + \beta_2\,(\textit{demographic factors}) + \beta_3\,(\textit{cognitive reserve factors}) + \beta_4\,(\textit{error})$$

Clinical variables were considered to be those that are caused by MS (such as MS type, disease duration, level of neurological disability, number of relapses, and taking disease modifying treatment (DMT)) and those related to MS but not necessarily caused by MS alone (levels of depressiveness and antidepressant uptake). MS-unrelated predictors include demographic characteristics (age and

gender) and cognitive reserve factors (involving educational exposure, pre-morbid IQ and premorbid cognitive leisure, number of languages spoken, engagement in physical fitness, and employment status). The total amount of variance left unexplained by such a model would include measurement error, genetic variability in CNS plasticity / repair capacity, and potentially other unspecified moderator variables.

It is predicted that this model (or some of its components) would explain why some individuals experience more longitudinal changes in cognition than others.

1.6.4. Predictions for self-perception of cognitive change

After having measured the extent, dimensionality and causes of longitudinal changes in cognitive functioning, the third aim of the project was to investigate to what extent the patients themselves are aware of their deficit progression. Based on the analysis of previous literature I raised the assumption that pwMS may not be entirely accurate in their evaluations of their cognitive performance, and if that's true, I wanted to know what causes the patients to perceive the changes in cognition.
Based on the review of previous literature, the self-perception of longitudinal change in cognition could potentially be explained by the following model:

$$\textit{Self-perception of longitudinal change} = \alpha + \beta_1 \textit{(executive functioning)} + \beta_2 \textit{(depressive symptomology)} + \beta_3 \textit{(neurological disability)} + \beta_4 \textit{(Number of relapses in-between assessments)} + \beta_5 \textit{(MS impact)}$$

When building this model it was considered that the patients' perception of longitudinal change in cognitive functioning could be influenced by their abilities to make judgements (executive functioning), whether they can be identified as being more depressed, or whether they feel more affected by MS (either having a higher neurological disability score, a higher MS impact score or experiencing more relapses).

However, since heterogeneity in cognitive domains affected is expected, it can be that this model may have a better fit for self-perception of changes in cognitive functioning for some cognitive domains but not the others.

1.6.5. Concluding remarks

With this project I hope that by testing these predictions I will help fill the indicated gaps of knowledge. In this project both the factors that have positive and detrimental effects on cognition will be accounted for, and it is hoped that these analyses will be able to account for the individuality and heterogeneity in prevalence and rate of progression of cognitive impairment in MS.

The analyses carried out throughout this thesis were devised to combine the findings from previous studies into one multidimensional model that will potentially encompass and help explain the heterogeneity of trajectories of MS, and this way will add benefit the existing body of literature. Besides the theoretical value, throughout this project I also attempt to critically evaluate the existing methods of cognitive assessment and result analysis, therefore the findings from this project have the potential to benefit work in everyday clinical practice.

1.7. Structure of this thesis

This thesis begins with a presentation of the methodology of the phase I (baseline) assessment (Chapter Two). Chapter Three contains characterization of the phase I study, and the procedure behind the optimization of the phase II (follow-up) battery. Chapter Four contains characteristics of the phase II cohort. Chapter Five includes the analyses of the dimensionality of cognitive impairment in MS. Chapter Six addresses the trajectory, extent, and predictors of cognitive change in pwMS. In Chapter Seven I analyse the patients' self-perceptions of cognitive change and compare them to the actual change in performance.

Since the majority of the experimental chapters included overlapping themes (definition of impairment, comparison of patients to controls, and clinical implications) to avoid redundancy all results were discussed together in the final discussion chapter (Chapter Eight).

Chapter Two. Phase I battery

2.1. Chapter overview

This chapter describes the methodology of the phase I of the longitudinal study. It starts with a description of the design and the recruitment procedure, followed by a detailed presentation of the materials and the administration procedure. The chapter ends with an outline of the strengths and weaknesses of the phase I test battery.

2.2. Phase I design

The phase I study was started as a PhD research project at The University of Edinburgh by Mara Sittampalam in 2010 and was supervised by Dr Thomas Bak and Prof Siddharthan Chandran. The phase I study had a cross-sectional between groups design and its aim was to gain a better understanding of the prevalence and underlying mechanisms of cognitive dysfunction in MS, and then to relate them to MS subtype, disease duration and motor disability scales.

2.3. Phase I study recruitment

For the phase I study the patient participants were recruited from secondary care in the Lothian area of Scotland, referred by local neurologists and specialist nurses between August 2010 and August 2012. Potential participants were screened against the following eligibility criteria: revised (2010) MacDonald criteria MS (Polman, et al., 2011), ages 18-65 years inclusive and absence of psychiatric or physical comorbidity (including major affective disorder, significant dementia or other significant comorbidities). Subjects showing any non-MS related ophthalmological condition that might interfere with neuropsychological testing were excluded too. For this purpose subjects having a Snellen acuity worse than 20/70 (0.25) (criterion for visuoperceptual testing proposed by McCarthy and Warrington (McCarthy & Warrington, 1990)) were excluded. Furthermore, it was chosen to restrict the phase I study to MS patients not residing permanently in a nursing home or other institutional setting. Of 972 pwMS screened, 108 patients fulfilled the eligibility criteria and agreed to participate. It was projected to accomplish a sample with the prevalence of PPMS, SPMS and

RRMS similar to the natural prevalence in the population. Based on that at phase one Mara Sittampalam has recruited 21 patients with PPMS, 34 SPMS and 53 RRMS totalling to 108 people in the MS sample.

To enable definition of appropriate normal ranges a matched control cohort was then achieved through stratified sampling by age, gender and educational level. Exclusion criteria were a disease of the central nervous system, major psychiatric illness, history of alcohol or drug abuse, serious head injury, learning disability and recent heart attack or other major medical illness. All persons needed to have adequate vision to complete testing, and if participants required hearing or visual aid, they were used during all assessments. A parallel group of 33 healthy controls included patients' spouses, family members and friends, as well as selection from the Psychology Department of The Edinburgh University volunteer panel.

2.4. Phase I study methods

This section includes a description of the tools employed to collect data on demographic, disease, disability and neuropsychological variables. The information about the validity of the assessment tools was reported only for the items that were included in the optimized phase II battery, and that information can be found in Appendix D.

2.4.1. Demographic and disease information

The demographic data collected about participants included age, gender and years of education. For the patient participants the history of MS, including, diagnostic criteria, disease onset and disease duration, medication, motor and sensory symptoms past and present relevant to Kurtzke disability score in MS Expanded Disability Status Scale (EDSS) was collected.

2.4.2. Assessments of disability

It has been shown that disability has a large impact on ability to perform certain cognitive tasks. This may be due to the nature of standard cognitive batteries that are problematic for patients with speech or upper limb weakness. Even though this was not considered to be a problem with the majority of tests being used in this study, physical disability was measured to control for this confounder.

EDSS: The Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) combines an assessment of seven different neurological domains and an ambulatory score. These domains are referred to as Functional Systems (FS) and include pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, and cerebral (or mental) functions. The patient's experiences of symptoms are not included, merely the findings from physical examination and direct questions related to mood, cognitive and bladder and bowel symptoms. The EDSS scale is 20 half steps from '0' to '10', with '0' corresponding to a completely normal examination, and '10' to death due to MS. The scores from '0' to '4' depend on individual FS scores, and over '4' the score is based on assessment of ambulation using distance and dependence on aids. Since its introduction, it has become standard practice to use the EDSS score in MS research and clinical assessments. As part of EDSS visual domain visual acuity data for right and left eyes and binocular vision with correction was collected using a 2m Snellen acuity chart in order to exclude participants with visual problems that might interfere with neuropsychological evaluations.

MSIS 29: The Multiple Sclerosis Impact Scale (MSIS-29) (Hobart, Lamping, Fitzpatrick, Riazi, & Thompson, 2001) comprises of 20 physical and 9 psychological questions based on a 1-5 Likert scale. Unlike other patient based questionnaires, MSIS-29 was developed using psychometric methods. 129 items gained from patient interviews, literature review and expert opinion, were then systematically reduced after analysing protocols from 1530 patients, and a 29-item scale was developed. Item descriptive statistics, item convergent and discriminant validity, and factor analysis indicated that it was legitimate to generate scores for MSIS-29 scales by summing items. These results indicate the MSIS-29 is a clinically useful and scientifically sound patient-based outcome measure of the impact of MS suitable for clinical trials and epidemiological studies.

BDI - II: Beck's Depression Inventory – II (BDI - II) (Beck, Steer, & Brown, 1996) is a questionnaire that is used in research and clinical practice, in a wide range of medical diseases and is composed of items relating to symptoms of depression such as hopelessness and irritability, feelings of guilt or being punished, as well as physical symptoms such as fatigue, weight loss, and lack of interest in sex. The scale takes 5 minutes to complete and can be self-administered or verbally by a trained administrator. BDI-II contains 21 questions, each answer being scored on a scale value of 0 to 3.

Higher total scores indicate more severe depressive symptoms. The test was also shown to be robust to daily variations in mood (Beck et al., 1996).

2.4.3. Assessments of general cognitive ability

ACE-R: Addenbrooke's Cognitive Examination – Revised (ACE-R) (Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006) is a screening battery for dementia that incorporates the Mini-Mental State Examination (MMSE) with the addition of verbal fluency, visuospatial tasks and expansion of the memory and language components. It is a brief (takes around 15 minutes) bedside screening tool that is easy to conduct and score in routine clinical practice (Larner, 2007). The ACE-R it is widely used across the world, having been translated into over 30 languages and validated in different cultures (Bak & Mioshi, 2007). Importantly, it has been applied in a broad range of neurodegenerative diseases, including dementias (Mathuranath, 2000) and neurodegenerative movement disorders (Bak et al., 2005). The ACE-R assesses five cognitive domains, and each of their scores add to a composite score of 100: attention (18), verbal fluency (14), language (26), memory (26), and visuospatial skills (16). The ACE-R score ranges from 0 to 100, with higher values indicating better cognitive abilities. A score of 88 is reported to be a sensitive cut-off for detecting mild cognitive impairment and a score of 82 is an indicator of dementia (Mioshi, et al., 2006).

NART–2: National Adult Reading Test, second edition (NART-2) (Nelson & Willison, 1991) is a test designed to provide an estimate of premorbid cognitive ability. The subjects are required to read a list of 50 irregularly spelled words (e.g. heir, banal, zealot) and assuming that the patient is familiar with the irregular word, accuracy of pronunciation is used to predict IQ, and is argued to depend more on previous knowledge than on current cognitive capacity, as the ability to pronounce irregular words is generally retained in mildly demented individuals (Stebbins, Wilson, Gilley, Bernard, & Fox, 1990). The NART-2 reading scores can further be used to predict pre-morbid WAIS-R IQ. In this study the predicted WAIS-R IQ scores (accounting for demographic differences) were used instead of raw NART scores in all analyses, but for the purposes of convenience they were indicated as 'NART'.

2.4.4. Assessments of memory, attention, processing speed and verbal fluency

BRB-N: Brief Repeatable Battery of Neuropsychological tests (Rao, 1990) is a sensitive measure of cognitive impairment in pwMS. BRBN provides measures of sustained attention/concentration, information processing speed, verbal and visuospatial learning and delayed recall and semantic retrieval (fluency). Such functions are reported to be most often disrupted in pwMS (Peyser et al., 1990). Administration of the total test battery takes about 20-30 mins and it consists of the Selective Reminding Test (SRT) (8mins), the 10/36 Spatial Recall Test (SPART) (5mins), the Symbol Digit Modalities Test (SDMT) (3mins), the Paced Auditory Serial Addition Test (PASAT) (10mins) and the Word List Generation Test (WLGT) (2mins). The test battery is administered in the following order: SRT, SPART, SDMT, PASAT, Delayed Recall of the SRT, Delayed Recall of the SPART and the WLG.

The Selective Reminding Test (SRT) is a test used to measure verbal learning and memory during a list learning task of six trials. The list consists of 12 words which the examiner reads at a rate of one word per two seconds. The subject is instructed to recall all 12 words in any order. Every consecutive trial only the words that are missed on the preceding trial are given. After 15 min (following the administration of the PASAT) the subject is asked to recall the word list. The SRT distinguishes between short-term and long-term components of memory and examines also the consistency of retrieval from long-term memory. The scoring is according to published rules (Buschke & Fuld, 1974). Three SRT indices were used in our study. A word recalled on two consecutive trials is considered to have entered long-term storage (LTS) on the first of these trials and scored as LTS on all following trials regardless of subsequent recall. The total sum of the words in LTS of all six trials is taken (SRTL). If a word in LTS is consistently recalled on all subsequent trials then it is scored as in Consistent Long Term Retrieval (CLTR). The total sum of the words in CLTR of all six trials is taken (SRTC). As per Rao's instruction manual (Rao, 1990) we used corrected SRT LTS and SRT CLTR values for every male participant, where we added four points to their scores. The Delayed Recall (SRTD) is the total number of words recalled after the delayed period.

The 10/36 Spatial Recall Test (SPART) was developed to assess visuospatial learning and delayed recall. The test consists of a 6X6 checkerboard with ten checkers randomly placed, this is an adapted

version of the 24X7 Barbizet memory test (Barbizet & Cany, 1967). The board is put in front of the subject for 10 seconds. After presentation the subjects attempt to reproduce the original design on an empty board. This process is repeated twice and after 15 minutes (following the SRTD) the subject is asked to recall the design again. The score is the total number of correct responses for the three trials (SPART) and the delayed recall trial (SPARTD).

The Symbol Digit Modalities Test (SDMT) examines information processing speed by primarily assessing complex visual scanning and tracking (Smith, 2002). The subject examines a series of nine meaningless geometric symbols which are labelled 1 to 9. During 90 seconds the subject substitutes symbols in a row by the corresponding number and responds verbally. The score is the number of correct substitutions (SDMT).

The Paced Auditory Serial Addition Test (PASAT) is a measure of sustained auditory attention and information processing speed (Gronwall, 1977). The subject is instructed to add 60 pairs of digits such that each number is added to the one that immediately precedes it and report the outcome verbally. The digits are presented by tape, with a speed of every 3 seconds one digit. The subject is required to respond verbally quickly, inhibit encoding of his own response while attending to the next stimulus in a series, and perform at an externally determined pace. The score is the number of correct responses per trial (PASAT).

The Word List Generation (WLG) is a semantic verbal fluency test evaluating the spontaneous production of names of a given category within a limited amount of time (Rao, 1990). The subject is asked to give as many names of vegetables and fruits (version A) or animals (version B) as possible during the time given. The score is the number of correct names produced (WLGT). Since we administered version B, to avoid redundancy, we used the animal fluency scores from the ACE-R battery. That task took 60 seconds to complete and we used the scoring norms as in the ACE-R administration protocol (Mioshi et al., 2006).

Operational definition for failing the BRBN battery was chosen to be performance below cut-off for at least two of the eight BRBN subtests as described elsewhere (Amato et al., 2010; Borghi et al., 2013).

2.4.5. Assessments of visuoperceptual functions

All of the stimuli in the visuoperceptual tests were administered binocularly, printed in black and white and the assessments were not timed. The visuoperceptual assessment was restricted to operations on non-moving stimulus configurations, and tasks requiring rapid verbal or motor responses were avoided. The subjects who were uncertain were urged to guess.

VOSP: Visual Object and Space and Perception battery (VOSP) (Warrington & James, 1991) is a collection of eight tests based on the theory proposed by the authors that object and space perception is functionally and anatomically independent. Each of the eight VOSP test was devised to focus on one component of visual perception, while minimising the involvement of other cognitive skills. The VOSP consists of four tests of object perception (VOSP1, VOSP2, VOSP3 and VOSP4) and four tests of space perception (VOSP5, VOSP6, VOSP7 and VOSP8). In the phase I of this study we only used VOSP2, VOSP3, VOSP5, VOSP6, VOSP7 and VOSP8.

In the Silhouette test (VOSP2) the subject is shown an outline of an object and asked to identify by naming them. The test consists of 15 silhouette drawings of animals and 15 silhouette drawings of inanimate objects and was constructed to be of graded difficulty ranging from very easy silhouettes that could be identified by all subjects to difficult silhouettes that only a proportion of the control sample identified.

The Object Decision (VOSP3) test requires the participants to look at four figures of which one is a silhouette of an object and other three are distractor items, object-like shapes that are entirely imaginary, and to identify which one of them is a silhouette of a real object.

In the Dot Counting Test (VOSP5) the test stimuli consists of arrays of black dots on a white card and the subject is asked to count them.

The Position Discrimination test (VOSP6) stimulus consists of two adjacent horizontal squares, one with a black dot printed exactly in the centre and one with a black dot just ‘off’ centre. The subjects are asked to identify the dot which is in the centre of the square.

In the Number Location task (VOSP7) the subjects are presented with two squares, the top one contains randomly placed numbers (1-9) and the bottom square contains a single black dot corresponding to the position of one of the numbers. The task is to identify the number that corresponds with the position of the dot.

In the Cube Analysis test (VOSP8) the participants are presented with drawings of 3D arrangements of square bricks and are asked to count how many solid bricks are represented in each of the drawings. This test provides a measure of the perception of complex spatial relationships.

BORB: Birmingham Object Recognition Battery (BORB) (Riddoch & Humphreys, 1993) consists of 14 separate subtests designed to assess particular aspects of visual processing and visual object recognition. In this study only the four BORB subtests dealing with the processing of 'pre-categorical' properties of objects (i.e. properties not tied to stored knowledge about the particular objects involved) were included. These tests are directed solely at the perception of basic properties and object forms – those include their size, orientation, location and length. In these four tests two items per trial are presented. The items are either the "same" (i.e. match in length, size, etc.) or they are different. There are equal numbers of "same" and "different" trials which are mixed together randomly. The patient is to indicate which stimuli are the same and which are different. Test 2 (BORB2) requires the matching of the line length, Test 3 (BORB3) the matching of the stimulus size, Test 4 (BORB4) the matching of line orientation and Test 5 (BORB5) the matching of the positions of gaps in two circles.

2.4.6. Assessments of language functions

Tests were chosen with consideration to the abilities and challenges of people with MS, therefore none of the tests chosen are timed, and the physical strain was kept to a minimum, with limited written and oral responses (many of the tasks involve only pointing). All computer-based tests were presented using a Toshiba laptop computer with a 10-inch touchscreen.

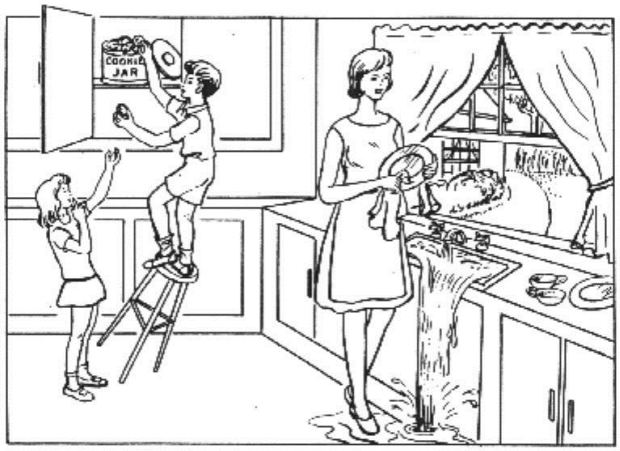
TROG: Test of Reception of Grammar (TROG) (Bishop, 1982) is a receptive language test, designed to assess understanding of English grammatical contrasts marked by inflections, function words and word order. In this test the participants hear an orally presented sentence, then point to one of four

pictures which illustrates the content of the sentence. Additionally, this test has the practical advantage of being widely used and standardized in both child and, increasingly, brain damaged populations, and is frequently used to assess adults with aphasia. TROG enables the tester to discover not only how a person's grammatical comprehension compares with that of other people of the same age, but also to pinpoint specific areas of difficulty. In order to focus on the syntactic properties of the examined structures, the vocabulary utilized and the visual form of the pictures is kept as simple as possible. In addition, the sentence length is kept constant for tasks of different syntactic complexity, minimizing the effect of possible attentional and short-term memory impairment. These two aspects are of particular importance when investigating patients with neurodegenerative diseases affecting several cognitive domains (Croot, Hodges, & Patterson, 1999). A multiple choice test requiring no verbal output, the original version comprises of 80 test items divided into 20 blocks of four test items with each block testing a particular grammatical construct, arranged in order of increasing complexity. In order to decrease the time taken, only the second 40 items of the tests were used, blocks K to T, as they were shown to be more sensitive to impairment in neurodegenerative diseases (Bak, O'Donovan, Xuereb, Boniface, & Hodges, 2001) and enabled us to shorten the overall testing time.

All sentences were presented to the participants regardless of previous correct or incorrect responses, and each question was scored according to pass/fail system, with a maximum score of 40.

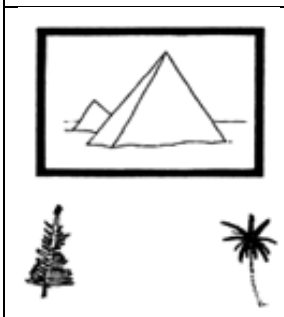
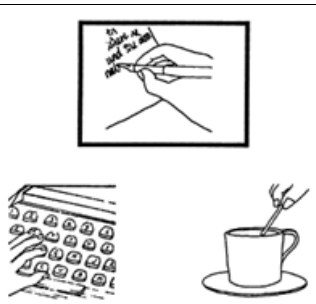
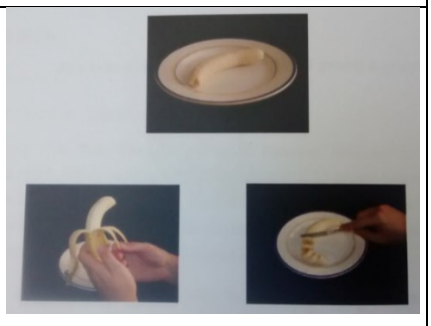
BCT: Boston Cookie Theft test (BCT) (Goodglass & Kaplan, 1983) is a picture description and spontaneous language production task that measures expressive syntax. The participants are shown a picture of an eventful family kitchen scene that contains many details of surroundings familiar to most people and may be described with a simple vocabulary acquired early in life (Giles, Patterson, & Hodges, 1996) (Figure 2.1). The participants were asked to write their descriptions down rather than saying them aloud in order to promote grammatical production, and only for those patients who had severe upper limb disabilities or tremor the test was administered orally. The text was analysed according to a protocol developed by Friedel M. Reischies research group in Berlin (Bschor, Kühl, & Reischies, 2001), where the number of words and the number of picture variables (correctly named persons and objects, localizations, actions and features) was interpreted.

Figure 2.1. Boston Cookie Theft test stimulus. BCT test assesses spontaneous speech as the participants are asked to describe the scene from the stimulus picture



PPT, KDT and TTT: are computer generated tasks that assess different aspects of understanding of semantic relationships. The test employed for receptive semantics was a shortened version of the Pyramids and Palm Trees Test (PPT) (Howard & Patterson, 1992), the Kissing and Dancing Test (KDT) (Bak & Hodges, 2003), and the Tomatoes and Tuna Test (TTT) (Danek et al., 2013). The PPT is concerned with the relationships between nouns, the KDT deals with verbs, and the TTT, sequential events and cause-effect scenarios. All of these tests involve making decisions about sets of pictures (Figure 2.2).

Figure 2.2. An example of a word and picture triplets from the three semantic tasks.

PPT	KDT	TTT
		

Abbreviations: PPT – Pyramids and palm trees test; KDT – Kissing and dancing test; TTT –Tomato and tuna test.

Participants must choose which of two alternatives at the bottom is best associated with the image at the top. For instance, in PPT, which assesses perceived relationships between nouns, palm trees are

associated with pyramids according to geography, but the other option, pine trees, has no such connection. In KDT, which assesses perceived relationships between verbs, writing is associated with typing, but not stirring coffee. In TTT, which assesses sequential events and cause-effect scenarios, the top picture of peeled banana is the goal state, and only the bottom picture of peeling the banana will be the action preceding the goal state and therefore correct.

In this study we used the abbreviated computerised version of these three tests where each test contains 25 tasks of similar difficulty (Hulst, 2012). The tests required only touching the correct image on the computer screen and were simple and straightforward to perform. These tests did not require any verbal response and thus could be regarded as a more 'pure' measure of semantic function than, for example, regular picture naming tasks.

GNT: The Graded Naming Test (GNT) (McKenna & Warrington, 1983) is a naming test that assesses expressive semantics and was chosen due to its popularity and the range of difficulty it provides. Given the variability other researchers have observed in naming tests, it was important to prepare for the possibility of different levels of naming ability. This confrontation naming test comprises of 30 black and white drawings with items decreasing in frequency from item 1 'kangaroo', to item 30 'retort'. Items were presented one at a time and participants were asked to give the name for each item. Responses were marked as correct only if named exactly as the target, except for item 25 'yashmak' where 'hijab' was also accepted.

Minimal Pairs Non-Words: Minimal pairs non-words (MPNWDS) discrimination task is a test of receptive phonology. Words that differ by only one phoneme are spoken aloud by recorded voices in a powerpoint slideshow developed by Philippa Jane Rewaj (Rewaj, 2013). In the minimal pairs test, participants needed to respond if the words they hear are the same or different (e.g. deg – ged). The nonword test comprised of 48 pairs, with 24 'same' pairs and 24 'different' pairs. There were two voices in the recorded stimulus, one male and one female – from the East and West of Scotland, respectively – but both words of a given minimal pair were spoken by the same speaker. There was a one-second delay between the words, and participants were asked to give their judgement aloud immediately after hearing the stimuli. The non-words were of comparable length and syllabic

complexity to real English words, never violating the rules of English phonology (eg. tusset, tuzzet). The phoneme that differed between the two words was always a consonant, and would vary only by one feature (place, manner or voicing).

2.4.7. Assessments of audio-visual integration

The Orchard task: The Orchard task is a relatively difficult test of audio-visual non-semantic information processing that was developed by Professor John Hodges and Dr Thomas Bak in 2007 in Cambridge. During this task the participants are presented with a string of tones followed by four pictures with different numbers of dots inside the squares. The participants are required to listen to the recording and count how many tones they hear to determine which one of the four dot pictures matches the number of tones they have heard. This test is comprised of ten trials that are arranged in increasing difficulty.

Sound picture matching: Sound picture matching task (SPMT) is a test developed for testing the ability to integrate auditory and semantic visual information (Hulst, 2012). Unlike the Orchard task, the SPMT requires access to previous knowledge of how everyday objects and animals sound. The participants were presented with a sound, for example, a recording of a song played on the piano, and were shown four pictures of musical instruments, namely French horn, piano, harmonica and a saxophone. Then the participants were asked to identify which one of the four objects could be associated with that sound. The test included 30 trials starting of increasing difficulty, and the sounds represented everyday man-made objects and living animals.

2.5. Phase I procedure

To minimise the effect of fatigue, at phase I the neuropsychological testing was conducted over at least two sessions where the patients were asked to come in to undergo clinical and neuropsychological assessments. In combination both visits took at least four hours to complete, with additional sessions being scheduled for participants who couldn't manage to go through the tests in the designated time, or for participants that got tired quickly and it was decided to continue over the third visit a few weeks later. For those participants the total testing time could go up to six

hours to complete. The order of the assessments was that during the first visit the clinical information was collected first, followed by tests of general cognition, memory, attention, processing speed and verbal fluency, and in the second visit tests of visuoperceptual cognition and language were administered in the same order as they have been presented in this chapter.

2.6. Phase I battery strengths and weaknesses

The test battery developed by Mara Sittampalam has been devised to be a comprehensive set of tools designed to pick up and identify domain-specific MS-related cognitive impairment. To our knowledge this investigation conducted by Mara Sittampalam has been the most in-depth and detailed analysis of multi-domain MS-related cognitive impairment. No other study to date has assessed so many cognitive functions in such great detail on a large number of MS patients. Therefore with all this information collected on the clinical history, cognitive and physical abilities, and demographical information, this study contained vast amounts of information about possible predictors of MS-related disease progression and cognitive decline, thus formed a rich foundation for future longitudinal follow-ups. To our knowledge, to date no other longitudinal study including pwMS had such a vast baseline assessment incorporating multi-domain cognitive data which could be analysed and used in follow-up assessments as an indicator or biomarker of future MS-related cognitive decline. Therefore the work conducted at phase I represented a great foundation for a multi-domain longitudinal study of cognitive changes in MS.

Although the test battery presented above had superior coverage to any of the assessments in MS to-date, it was considered to be substantially too long. At the time between the phase I and phase II assessments some of the patients' MS had substantially progressed and five to six hours over three days would become unbearable for them to endure. Therefore for this reason only, the otherwise willing participants had the potential to end up declining the offer to participate. Since one of the most important aspects indicating the quality of a longitudinal study is its follow-up recruitment rate, we aimed to make sure that everything possible was done to ensure that high turn-up rates at the follow-up assessment were obtained. Thus one of the main objectives of the phase II protocol development was to create the test battery as short as possible so that more participants would be willing to come back; but at the same time to retain its informativeness and wide coverage of

cognitive domains. The protocol behind the optimization procedure of the phase II battery is presented in the next chapter, Chapter Three.

2.7. Chapter summary

In Chapter Two the methodology behind the data collection for the phase I assessment was presented. In this chapter the design, recruitment, materials and administration procedure of the phase I study were covered.

The phase I study had a cross-sectional between-groups design comparing the cognitive performance of pwMS with that of healthy controls. The aim of the phase I study was to gain a better understanding of the prevalence and underlying mechanisms of cognitive dysfunction in MS, and then to relate them to MS subtype, disease duration and motor disability scales.

During the phase I study 108 pwMS and 33 healthy gender and education-matched controls have been recruited. The participants have underwent extensive cognitive and clinical examinations that took four to six hours to complete and had been administered through two to three research visits. In this chapter the assessment battery employed for the phase I study has been presented in detail including the information collected regarding the demographic and disease variables, measures of disability (EDSS, MSIS 29, BDI-II), assessments of general cognitive ability (ACE-R, NART), memory, attention, processing speed and verbal fluency (BRBN), visuoperceptual functions (VOSP, BORB), language functions (TROG, BCT, PPT, KDT, TTT, GNT, Minimal Pairs Non-Words), and audio-visual integration (The Orchard task, Sound Picture Matching).

The test battery developed for phase I assessment has been devised to be a comprehensive set of tools designed to pick up and identify domain-specific cognitive impairment. No study to date had assessed so many cognitive functions in such great detail on a large number of MS patients, therefore the phase I study contained vast amounts of information regarding possible predictors of MS-related disease progression and cognitive decline, forming a rich foundation for longitudinal follow-ups. However, in order to proceed with reformatting the study design into a longitudinal cohort study, it was concluded that the necessary optimizations to the assessment battery were needed to be performed.

Chapter Three. Protocol development for the phase II study

3.1. Chapter overview

This chapter describes the theoretical and practical considerations that influenced the development of the phase II protocol. Standard study design issues were addressed, and a detailed account was given of revisions to the phase I cognitive assessment battery to ensure optimisation with respect to the phase II study aims.

3.2. Rationale for the development of a bespoke phase II protocol

The overarching principal for phase II protocol development was to ensure that the longitudinal study aims could be addressed. The ability to make valid linkage and comparisons between the phase I and phase II datasets was therefore essential. This requirement could be superficially achieved through simple repetition of the phase I study protocol. However, revision of the phase I protocol was necessary due to:

- International ethical standards for medical research involving human subjects as outlined in the World Medical Association (WMA) Declaration of Helsinki (WMA, 2013).

  "The design and performance of each research study involving human subjects must be clearly described and justified in a research protocol."

  Given the differing objectives of the phase I and phase II studies, *independent* development and justification of the phase II assessment schedule was necessary. In particular, justification of the phase II study assessment schedule required identification and inclusion of the metrics from phase I that were relevant and necessary to answer the specific questions being addressed at phase II, and removal of those that were irrelevant and/or unnecessary.

- Pragmatic considerations identified through feedback from phase I participants on the tolerability of the assessment schedule. The participants have found the phase I study to be too long and an unchanged assessment schedule was likely to impact negatively on recruitment to phase II.
- Advances in the understanding of cognitive impairment due to MS that have emerged since the phase I study was designed in 2010.

3.3. Phase II study design – general considerations

The phase I study was designed by a PhD student Mara Sittampalam and her supervisor Thomas Bak to be an independent piece of observational research. Phase I was therefore configured as a cross-sectional study without specific consideration to the possibility of longitudinal follow-up. An initial challenge was therefore to design a protocol that would enable conversion to a longitudinal study design, with phase II representing the first 'wave' of review and enabling the potential for subsequent 'waves' as independent follow-on projects.

3.3.1. Study aims

This study was designed to answer the research questions raised in the introduction and to evaluate the predictions made based on the review of previous literature.

The first aim was to evaluate the dimensionality of cognitive impairment in MS. We questioned the popular view that MS affects only frontal functions: complex attention, information processing speed, (episodic) memory, and executive functions (Jongen, Ter Horst, & Brands, 2012). We postulated that an underappreciated subgroup of patients exists in whom pathology affecting 'posterior' brain regions results in detectable impairment(s) of posterior cognitive functions such as visuoperceptual and language abilities. We parsed this as two linked but separate questions: (i) *what is the dimensionality of impairments in terms of cognitive abilities (test performance)*, and (ii) *what is the dimensionality of our cohort with respect to recognisable profiles (clusters) of impairment* (*i.e.* are there subgroups of the population with recognisable impairment profiles). These questions are separately addressed with respect to the phase I and phase II datasets in *Chapter Five – The*

Dimensionality of Cognitive Impairment in MS. We predicted that the pattern of cognitive deficits would become more homogeneous over time at the population level (as individuals accumulate an increasing number of distinct cognitive deficits).

The second aim was to evaluate the predictors of longitudinal change in cognitive performance. We predicted that there would be a range of 'within subject, within domain' changes in the severity of cognitive impairment at the level of individual subjects – cognitive impairments could worsen, remain static or even improve over time; and that these possible trajectories could be seen within the same individual across cognitive domains. We then sought to explore the extent to which longitudinal changes in cognition could be explained by clinical, demographic and cognitive reserve variables. This aim is addressed in *Chapter Six - The Trajectory, Extent, and Predictors of Cognitive Change in People with MS*.

The third aim of the project was to determine whether the participants themselves were able to perceive the longitudinal changes in their abilities in each cognitive domain, and to explore the factors associated with self-perception of cognitive change. This is addressed in *Chapter Seven – Self-perceptions of Cognitive Impairment in MS*.

3.3.2. Study population

Consistent with a longitudinal study design, all phase II participants were drawn from the cohort who participated in the phase I study, comprising 108 pwMS and 33 healthy controls.

3.3.3. Study environment

All phase I assessments of participants with MS were conducted in a The University of Edinburgh hospital-setting either at the Western General Hospital, the Royal Infirmary Edinburgh, or at the Princess Alexandria Eye Pavilion, depending on the stage of the recruitment and the date of the patient availability. All assessments were performed in uniform settings that comprised of a well-lit quiet room with a desk. However, given the inevitable progression of disability during the interval between assessment visits, and to ensure maximal recruitment without biasing phase II towards inclusion of participants with minimal change in disability, it was decided to offer phase II

assessment visits at the participant's home. This was indicated to be a secondary option if the participant was unable / unwilling to be assessed in a hospital-setting.

The phase II hospital-based assessments for patient participants were conducted solely at the Anne Rowling Regenerative Neurology Clinic at a desk in a quiet and well-lit room, with a research nurse on-site during the assessments. The Anne Rowling Clinic is a University of Edinburgh facility at the Royal Infirmary Edinburgh that provides out-patient care for people with a neurological condition who have been referred by their GP or neurologist, and all clinical activity is undertaken in partnership with the UK's National Health Service (NHS). For assessments undertaken at the participant's home, the subjects were seen seated in a well-lit quiet room of their choice. The measures taken to ensure that these visits would capture comparable data to that collected at the Anne Rowling Clinic included removal of potential distractors (TV, radio, family members, pets etc) and administration of materials in the same order according to the study protocol.

Phase I and phase II assessments of healthy controls were conducted at the research laboratories of The University of Edinburgh Psychology Department. These rooms were identical to the rooms at the Anne Rowling Clinic, as they have been specially designed for cognitive assessments, and were quiet and well-lit and had only a desk with chairs inside.

To ensure comparability, all phase II assessments were administered by a single researcher (the author), in the exact same order as per study protocol. The influence of practice effects due to item familiarity was mitigated by employing alternate forms of the BRBN and ACE-R. Other tests did not have alternative versions available therefore had the same versions administered. Moreover, in order to control for any potential practice and other effects, when analysing the results the longitudinal changes in patient performance scores were standardized accounting for the variability of controls' scores (explained in detail in Chapter Six).

3.3.4. Time-interval between phase I and phase II assessments

In order to maximise recruitment at phase II, a pragmatic view was taken to allow participants flexibility in the precise time-interval between phase I and phase II assessment. Phase I data was

captured between February 2011 and October 2012, and phase II data was projected to be captured between July 2014 and June 2015. The anticipated interval between assessments was therefore approximately three years. This interval is long-enough to support a reasonable belief that any observed changes in cognitive performance were likely to be caused by disease progression rather than the short-term variability in function that typifies MS, noting also that cognitive deficits exhibited following a relapse can be reversible (Foong, et al., 1998).

3.3.5. Ascertainment and recruitment

Since the phase I study was designed to have a cross-sectional design, the phase I participants had not given prior consent for their details to be passed to additional investigators, nor to be re-contacted with respect to further research projects. These permissions were therefore sought as part of submissions for ethical approval. Permission was also sought to access medical records to screen against phase II eligibility criteria prior to sending invitations to participate. Invitations to participate were sent by post to those subjects fulfilling phase II eligibility criteria, and followed up after two weeks by a telephone call. Written informed consent was taken before any study-specific procedures were undertaken.

3.3.6. Eligibility criteria for phase II recruitment

- Participation in the phase I study
- Current residence in Scotland
- Not developed a medical or psychiatric disorder that would preclude comparable assessment of cognitive performance, and patients not currently being on a relapse
- Willing and able to provide written informed consent

3.4. Scope of the phase II cognitive assessment battery

3.4.1. Overview of optimization strategy for the phase II cognitive battery

Both empirical and theoretical justifications were employed to derive the optimised phase II neuropsychological test battery. Specifically, the following framework was applied:

I. Responsiveness of phase I tests to clinically relevant changes in cognition

I first identified the phase I tests with significant ceiling effects that would reduce their value in longitudinal research. I then evaluated the frequency of phase I test failure among pwMS compared to controls in order to identify those tests with evidence of sensitivity to MS-related cognitive difficulties. Although no tests were excluded purely on these criteria, this information was used to inform on test selection within a given cognitive domain.

II. Adequate coverage of cognitive domains

An empirical and data-driven approach (principal component analysis [PCA]) was used to evaluate the dimensionality of cognitive test performance, informing on the relationship between specific tests and the independent dimensions of MS-related cognitive impairment. This enabled identification of the most relevant tests from each principal component (dimension) to include into the follow-up battery. This approach established a core battery for coverage of all significant dimensions, while allowing to minimise redundancy (when appropriate) by ensuring that the participants didn't perform multiple cognitive tests that evaluate the same cognitive function. Recognising the relatively limited sample size at phase I, and the exploratory nature of such analyses, this empirical evidence was regarded as informative but subordinate to a separate evaluation of dimensionality and coverage using theory-based definitions of cognitive domains. Some tests were therefore included from within the same empirically defined dimension if theory-based models of cognition identified them as valuable measures of independent cognitive functions.

3.4.2. Responsiveness to clinically relevant changes in cognition

Temporal responsiveness to change could not be directly assessed from the cross-sectional phase I dataset. However, useful insights were possible by considering: (I) sensitivity to detect variability in cognitive function (*i.e.* the absence of a significant ceiling effect), and (II) sensitivity to detect impairment defined through a binary present/absent categorisation (*i.e.* sensitivity with respect to deficits that are likely to be *'clinically relevant'*).

I. Sensitivity to detect variability in cognitive function at phase I

Coefficients of variation (COV) were used to provide a standardised assessment of variability in phase I scores between cognitive tests with widely differing absolute ranges. The COVs were defined as the ratio of standardized deviation to the mean. Higher COVs represent a simultaneous assessment of variability attributable to the measurement error of each test (not independently assessed at phase I) and to 'true' variability in cognitive ability. The COV for pwMS and control performance at phase I can be seen in Table 3.1.

Table 3.1. Ranked coefficients of variation (COVs) in cognitive test performance of pwMS and controls

Test item	PwMS COVs	Control COVs
SRTC	69.56%	33.67%
PASAT	52.73%	31.09%
SRTD	51.16%	22.41%
SRTL	47.25%	22.63%
SPARTD	35.39%	25.93%
SPART	32.57%	23.11%
BC.Index	31.33%	11.64%
SDMT	30.83%	13.57%
WLGT	26.54%	18.88%
VOSP2	20.39%	13.73%
GNT	17.32%	14.75%
VOSP3	11.40%	7.58%
VOSP7	10.77%	4.09%
BORB5	9.11%	6.84%
BORB2	8.58%	5.62%
VOSP6	8.20%	2.15%
BORB4	8.18%	6.08%
TTT	8.15%	4.49%
BORB3	7.90%	6.39%
KDT	7.89%	6.98%
Orchard test	7.53%	4.45%
VOSP8	7.53%	3.08%
PPT	7.30%	6.51%
SPMT	7.18%	4.20%
MPNWDS	5.75%	2.14%
TROG	5.22%	3.79%
VOSP5	5.22%	1.86%

Table 3.1. shows COVs for pwMS and control participants' performance on the phase I cognitive tests. The COV was defined as the ratio of the standard deviation to the mean.

Abbreviations: pwMS – people with MS, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

As it can be seen from the Table 3.1, the tests with the highest COVs were the items from the BRBN, BCT and the WLGT tests. As was expected, we have identified the same tests that in the previous literature have been defined as most suitable to detect MS-related cognitive impairment. However, it is important to note that we found these tests to have high coefficients of variation for the control performance as well, suggesting that non-pathological variability and/or measurement error were significant. Therefore, in order to dissect the component attributable to true differences (variability) in cognitive function due to MS, the ratio of COVs was derived (pwMS / controls) and is presented in descending rank in Table 3.2.

As it can be seen from Table 3.2, the top six ranked tests were all visuoperceptual or language tasks. Although the COVs were small in absolute terms for these tests, the ratio suggested that a change in performance was highly likely to be attributable to MS.

Table 3.2. Ranked relative variability in cognitive test performance between pwMS and controls

Test	PwMS COV	Control COV	Ratio (pwMS COV/ control COV)
VOSP6	8.20%	2.15%	3.81
VOSP5	5.22%	1.86%	2.81
BC.Index	31.33%	11.64%	2.69
MPNWDS	5.75%	2.14%	2.69
VOSP7	10.77%	4.09%	2.63
VOSP8	7.53%	3.08%	2.44
SRTD	51.16%	22.41%	2.28
SDMT	30.83%	13.57%	2.27
SRTL	47.25%	22.63%	2.09
SRTC	69.56%	33.67%	2.07
TTT	8.15%	4.49%	1.82
SPMT	7.18%	4.20%	1.71
PASAT	52.73%	31.09%	1.70
Orchard test	7.53%	4.45%	1.69
BORB2	8.58%	5.62%	1.53
VOSP3	11.40%	7.58%	1.50
VOSP2	20.39%	13.73%	1.49
SPART	32.57%	23.11%	1.41
WLGT	26.54%	18.88%	1.41
TROG	5.22%	3.79%	1.38
SPARTD	35.39%	25.93%	1.36
BORB4	8.18%	6.08%	1.35
BORB5	9.11%	6.84%	1.33
BORB3	7.90%	6.39%	1.24
GNT	17.32%	14.75%	1.17
KDT	7.89%	6.98%	1.13
PPT	7.30%	6.51%	1.12

Table 3.2. Coefficients of variation (COV) are shown for pwMS and control participants' performance on the phase I cognitive tests. The COV was defined as the ratio of the standard deviation to the mean. The ratio of COVs is also shown, with ranking from highest (variability in the MS cohort scores greater than in controls) to lowest (variability in the MS cohort scores similar to that in controls).

Abbreviations: pwMS – people with MS, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

II. Detection of Cognitive Impairment at Phase I

In order to estimate the prevalence of cognitive impairment it was first necessary to define what a cognitive impairment is by establishing cut-off scores indicating a deviation from the norm.

a) Definition of Impairment on Neuropsychological Tests at Phase I

In order to define impairment on the cognitive tests used at phase I, the cut-off values from previously published test-validation studies were therefore reviewed. For the experimental tests which had not yet been validated and therefore lacked published cut-offs, we used the norms that the test authors have employed in their studies for other purposes. For cognitive tests where the authors suggest adjusting subject-specific thresholds based on age, gender, or years of education, these were reviewed separately. As it can be seen from the Appendix A, most of the cognitive tests included in the phase I battery did not have well-established cut-off values, and very few tests have had MS-specific normative values. Moreover, for several tests different studies suggest discrepant normative values.

Appendix A shows cut-offs at 1.5SD and 2SD below the validation control sample mean. In research concerning cognitive impairment in neurodegenerative diseases of older age it is a tradition to use a cut-off of 2SD or 5th centile below control performance (if the data's sufficiently normally distributed) but, however, since MS develops early in the lifespan, some authors suggest to employ higher cut-off values, such as 1.5SD below control mean (Amato et al., 2010; Borghi, et al., 2013).

For many of the neuropsychological tasks employed, the previously published normal ranges of scores were either based on different reference groups, or not available (e.g. BORB), or not published in detail (*e.g.* TROG and BCT test). Therefore to ensure that the cut-off values for all the phase I tests were available, contemporaneous, and comparably valid, definitions of impairment were independently derived using the phase I control data. Three cut-off values were considered for each test: performance below the 5th percentile, and two based on standard deviations below our control sample mean (1.5SD and 2SD). These cut-off values produced on the dataset from phase I control cohort can be seen in Table 3.3. These values were then compared to published normative values

(Appendix A) to select the most appropriate cut-off values for our sample that would in turn provide consistency and therefore interpretability in the context of the existing literature.

Table 3.3. Comparison of three common definitions for impairment on individual cognitive tasks. The 2SD, 5th centile and 1.5SD cut-off values were derived from the performance of the phase I control sample

	Test	2SD below mean value	5th centile value	1.5 SD below mean value
BRBN	SRTL	28	25	33
	SRTC	13	12	20
	SRTD	6	5	7
	SDMT	40	41	43
	PASAT	18	0	25
	SPART	12	14	14
	SPARTD	4	5	5
	WLGT	15	13	17
VOSP	VOSP2	17	17	19
	VOSP3	16	16	17
	VOSP5	10	10	10
	VOSP6	20	19	20
	VOSP7	10	9	10
	VOSP8	10	9	10
BORB	BORB2	24	24	25
	BORB3	25	25	26
	BORB4	24	25	25
	BORB5	32	32	34
	TROG	36	35	37
	BC.Index	32%	35%	35%
	PPT	21	21	22
	KDT	20	19	21
	TTT	23	22	23
	GNT	18	16	20
	MPNWDS	46	45	46
	Orchard test	9	9	10
	SPMT	27	26	27

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

After comparing the 5th centile, 1.5SD and 2SD below control mean cut-off values with the published norms (Appendix A), it was decided that the cut-offs at 5th centile and 2SD below control mean were most compatible to the published norms therefore most suitable to proceed with and allow comparisons of our results to previous work. To avoid the risk of having too liberal cut-offs that include too many control participants thus making Type 1 error, it was decided not to proceed with the 1.5.SD cut-off values. The main difference between 5th centile and 2SD cut-off values were the values on the PASAT test. Specifically, from our sample two of thirty-three controls (6%) had found the PASAT instructions too difficult to understand and thus had failed to perform even the practice items resulting in a PASAT score of 0. Table 3.3. shows that the 5th centile value for the PASAT is 0, but this problem could be avoided when using the cut-off value of 2SD.

Therefore the impairment on the tests in the baseline battery was defined to be at 2SD below control mean. It is important to note, however, that due to the reference group we used, in our sample the cut-off values chosen were higher for the majority of BRBN and VOSP tasks than those proposed by the test authors. Therefore as a consequence these tests became more sensitive (and less specific) to detect cognitive problems in the MS sample.

b) Defining normal performance at phase II

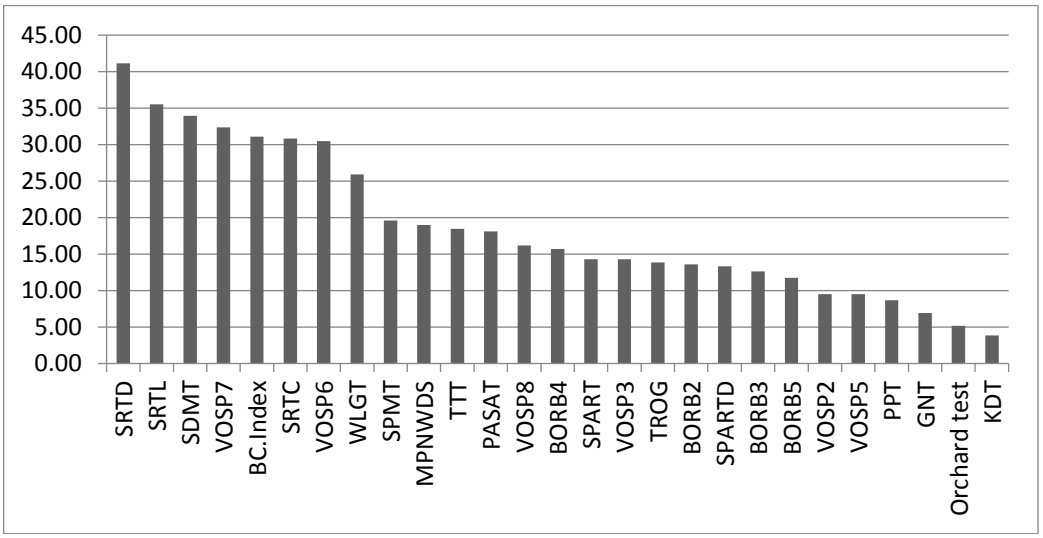
2 SD thresholds were also applied to the phase II control cohort and gave differing values for the majority of phase II tests. The differences between phase I and phase II cut-off values are presented in Chapter Five.

c) Frequency of impairment on neuropsychological tests

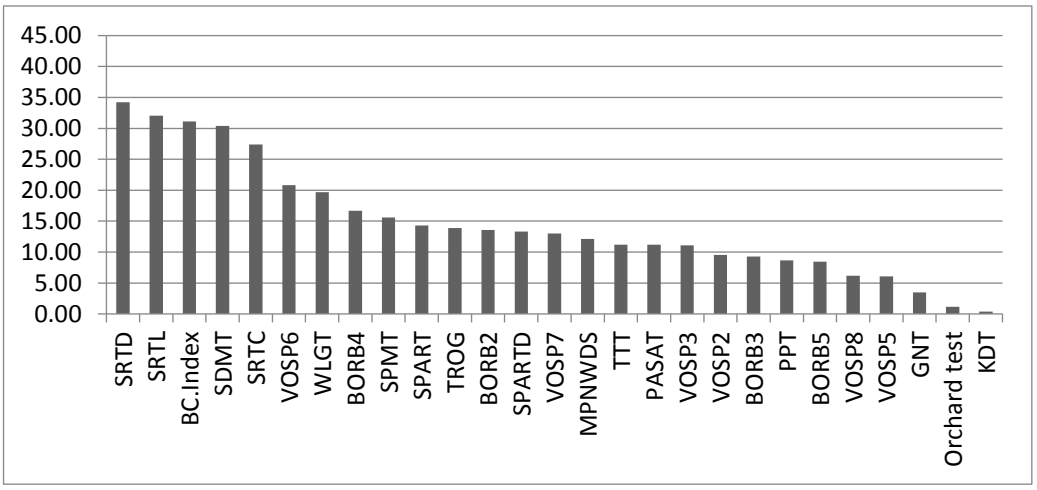
The frequency of cognitive test-failure for the phase I cohort is shown in Figure 3.1. These figures account for missing observations and the prevalence of impairment is expressed as a percentage (the exact number of collected observations for each test item can be seen in the table in Appendix B).

Figure 3.1. Frequency of cognitive impairment in pwMS by individual cognitive task at phase I

3.1 A. Raw frequency (%) of pwMS failing each cognitive task.



3.1 B. True frequencies (%) of pwMS failing each cognitive task (percentage of pwMS failing minus percentage of controls)



Abbreviations: pwMS – people with MS, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

Some tests had very few missing cases, for example the tests from the BRBN battery (105 to 107 of 108 observations), but some other tests had a lot of missing cases, such as the BCT test (90 of 108 observations). In the phase I assessment the missing data can be explained by which of the two or three study-appointments that particular individual has missed or at which point the fatigue became insurmountable. Missing data in the phase I dataset therefore could not be confirmed as missing at random, and under these circumstances imputation is unadvisable and thus was not performed. In the following result analyses the missing values were handled by exclusion.

As it can be seen from Figure 3.1A, more pwMS fail tests of memory and information processing speed than tests of semantic processing and naming. This has been a consistent and widely reported finding since Rao's seminal paper in 1991 (Rao et al., 1991).

It initially appeared that the rates of impairment were very similar across all BRBN tests that assess attention/concentration, verbal and visuospatial learning and recall, whereas there were widely differing rates of impairment on the visuoperceptual and language tests at baseline. To understand this, it was necessary to account for the relative difficulty of the tests. We investigated the true frequencies of MS patients failing each task by calculating how many controls have failed each of the tests and subtracting the frequencies of control failure (Figure 3.1B).

By investigating the true frequencies we have managed to identify the tasks that can be considered difficult to the patient participants only, from the tasks that are just challenging for everybody. This has helped in detecting the MS-related impairments.

3.4.3. Ensuring adequate coverage of cognitive domains and minimizing redundancy

I. Empirical evaluation of dimensionality in phase I cognitive performance

In order to evaluate the dimensionality of cognitive performance and the relationship of individual tests to identifiable dimensions, principal components analysis (PCA) was conducted using R-studio (version R 0.97.551) packages paran, princomp, GPArotation and rela.

Prior to running the PCA we checked whether our data met the relevant assumptions. First we checked each individual variable for normality of its distribution, and then separately we analysed to what extent the variables correlated with each other. Factorability was tested by analysing the sampling adequacy measures, and determining whether the sample size was sufficient.

a) Testing assumptions of data distribution

Prior to running the PCA the normality of distribution of test items was checked for using a Shapiro-Wilk test and by analysing the distribution histograms. The results of initial exploration of univariate normality can be seen in Appendix B.

The results from the initial exploration of normality with Shapiro-Wilk tests were discouraging. However, the Shapiro-Wilk test is often reported to be overly conservative, and individual histograms are typically considered to be more informative (see Appendix B for histograms for each variable). The normality of distributions of each groups of tests were discussed below:

BRBN items: SRTL, SRTC and SRTD, SDMT, SPART, SPARTD, and WLGT items were sufficiently normal to allow ongoing inclusion in analyses. The PASAT was markedly platykurtotic. Transformation could not alter this, but platykurtosis is considered to allow responsiveness to the latent variable, thus judgment was taken that it is preferable to include the PASAT scores.

Visuoperceptual items: All visuoperceptual items had significant negative-skew, which in this instance wasn't much of a problem since it meant that all visual cognition items were similarly easy to the participants. It could be argued that VOSP2 and VOSP3 and all of the BORB items were sufficiently normally distributed since they had only one peak and no clear ceiling effect. However, due to problematic negative skew (obvious ceiling effect) on items VOSP5, VOSP6, VOSP7 and VOSP8, I further explored whether the negative skew could be in part mitigated by logarithmic transformation.

Language items: The distribution of the language items faced similar problems as the visuoperceptual tests; they were also significantly negatively skewed. Only the BC.Index, PPT, KDT and GNT items could be argued to be not very different from normal distribution as they had only

one peak and less obvious ceiling effect. However, an obvious ceiling effect could be observed in the TROG, TTT, MPNWDS, SPMT and the Orchard tasks. It was therefore decided to attempt to logarithmically transform these items as well. Logarithmic transformations were employed for the non-normally distributed items, using the natural log (LN) function for each variable. As it can be seen from Table 2 in the Appendix B, the transformation did not improve the skewness of those variables.

For this purpose it was decided to plot the logarithmically transformed distributions of those variables and upon visual inspection to estimate the level to which the ceiling effect was prominent. The histograms of the logarithmically transformed distributions of these variables have also been included in the Appendix B. I have also attempted the following transformations: Log10(x), Sqrt(x), 1/x, where x was the variable to be transformed. I also tried running all of these transformations by replacing x with (k-x) where k was the highest observed value in each variable, but because the negative skew was so prominent, none of these attempts helped bring the distributions closer to normal. Since even after the transformation the ceiling effect was still very much prominent it was concluded that these variables had too much negative skew to be included into the PCA. Therefore it was decided to omit these items from the PCA.

The items included into the PCA were all eight BRBN items (SRTL, SRTC, SRTD, SPART, SPARTD, SDMT, PASAT, WLGT), six visuoperceptual tests (VOSP2, VOSP3, BORB2, BORB3, BORB4, BORB5) and four language items (GNT, PPT, KDT and BC,Index).

b) Testing assumptions of factorability

Correlations between variables: Since a large number of items were non-normally distributed (Appendix B) I chose to employ the non-parametric Spearman's test to measure the extent to which the items from the cognitive tests were inter-correlated (*i.e.* that there were at least some correlations amongst the variables so that coherent factors could be identified). All 18 items were entered to this analysis and I produced an 18x18 table of correlations (Table 3.4). From analysing the data presented in Table 3.4 it was concluded that the items were sufficiently inter-correlated to allow further analyses.

Table 3.4. Correlations among cognitive test items performed as part of principal component analysis. Phase I patients with full data (n = 95)

	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	VOSP2	VOSP3	BORB2	BORB3	BORB4	BORB5	GNT	KDT	PPT	WLGT	BCI.Index
SRTL																		
SRTC	**.888**																	
SRTD	**.699**	**.752**																
SPART	**.319**	**.339**	**.380**															
SPARTD	**.270**	**.280**	**.313**	**.721**														
SDMT	**.373**	**.412**	**.515**	**.425**	**.435**													
PASAT	**.318**	**.347**	**.253**	**.276**	**.293**	**.516**												
VOSP2	**.193**	**.198**	**.261**	**.274**	**.242**	**.364**	**.237**											
VOSP3	.156	.148	.065	-.03	-.04	**.186**	.165	**.353**										
BORB2	.079	.138	.110	.024	.077	.110	.108	.048	.020									
BORB3	.151	**.253**	**.286**	-.01	.030	.169	.084	**.310**	.125	**.418**								
BORB4	.132	.124	.163	.081	**.268**	.065	**.213**	**.218**	.042	.150	.134							
BORB5	.175	**.207**	**.194**	.103	**.185**	**.269**	.180	.161	.057	**.425**	**.254**	**.211**						
GNT	**.260**	**.284**	**.261**	**.242**	**.187**	**.307**	**.363**	**.505**	**.385**	.008	**.189**	.162	**.234**					
KDT	.155	**.245**	**.352**	.133	.124	**.294**	.084	**.303**	.061	.029	-.03	-.07	.143	.117				
PPT	.176	**.229**	**.335**	.034	.076	**.341**	.155	**.396**	**.238**	-.01	**.229**	**.200**	.165	**.367**	**.455**			
WLGT	**.327**	**.308**	**.344**	**.299**	**.369**	**.500**	**.437**	**.317**	**.254**	.110	.022	**.193**	**.226**	**.415**	.131	.135		
BC.Index	-.07	.023	-.04	.121	-.01	.075	-.06	.108	**.217**	-0.1	-0.1	-.04	-.01	.018	-.04	.053	.143	

The sections in **bold** represent statistically significant relationships ($p < 0.05$, unadjusted for multiple comparisons)

Abbreviations: SRT – Selective Reminding Test (SRL – Long Term Storage, SRTC– Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words.

Testing for sphericity: Bartlett test of sphericity was found to reach significance ($X^2(136) = 638.6$, $p < 0.001$), therefore it was concluded that the sample intercorrelation matrix did not come from a population in which the intercorrelation matrix is an identity matrix, and it was safe to proceed with further analyses.

Testing for sampling adequacy: I used the measures of sampling adequacy to investigate the degree of collinearity among the variables. The measures of sampling adequacy for individual items can be seen in Table 3.4. The BC.Index test item did not reach the minimum requirement for sampling adequacy and was therefore excluded from further analyses.

Table 3.5. Values of measures of sampling adequacy (MSA) for items considered to include into principal component analysis

Item	MSA
SRTL	0.694
SRTC	0.722
SRTD	0.840
SPART	0.691
SPARTD	0.680
SDMT	0.832
PASAT	0.800
WLGT	0.807
VOSP2	0.735
VOSP3	0.743
BORB2	0.570
BORB3	0.498
BORB4	0.488
BORB5	0.740
GNT	0.731
KDT	0.497
PPT	0.713
BC.Index	0.249
OVERALL MODEL	0.719

Abbreviations: SRT – Selective Reminding Test (SRTL– Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words)

For our sample matrix the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.719, which met the minimum criteria of being higher that 0.5 and it was reasonable to conclude that if a PCA was conducted, the factors extracted would account for a sufficient amount of variance but although not a substantial amount (Cerny & Kaiser, 1977).

Having removed all the missing observations, the following PCA was conducted on scores from 95 patients that had full data from phase I assessment on all 17 variables included into the PCA. Ideally for a PCA, there should be a large ratio of N / k (Cases / Items), e.g., > ~20:1, but PCA is considered to still be reasonably done with ~5:1, with bare minimum for pilot study purposes as low as ~3:1 (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Since there were 17 variables in the PCA and 95 patients with full data, this gave a ratio of 5.58, which was reasonably high to allow further investigations.

c) Number of retained components from principal component analysis

Scree plot: To determine the most appropriate number of factors to extract from the PCA I used Cattell's Scree Plot. This is a plot of Eigenvalues associated with each of the factors extracted, against each factor. Principal components with Eigenvalues higher than 1 are typically regarded to be worthy of retention due to their contribution of explanatory value. As shown in Figure 3.2, a four-factor solution was supported.

Figure 3.2. Scree plot to determine factor retention in principal component analysis

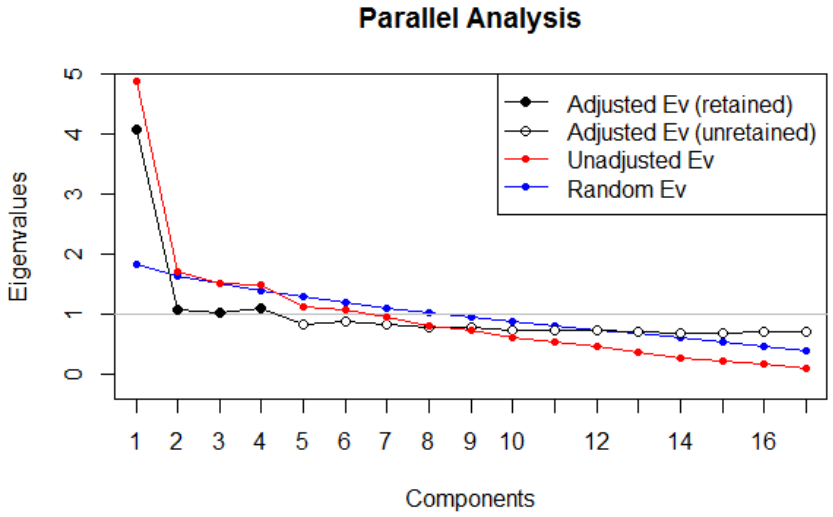


Table 3.6. Loadings from the first four factors from the Principal Component Analysis solution

Item	Factor 1	Factor 2	Factor 3	Factor 4	Communality	Uniqueness
SRTL	**0.93**	-0.03	0.00	-0.03	0.830	0.170
SRTC	**0.99**	-0.04	-0.01	0.03	0.964	0.036
SRTD	**0.73**	0.14	0.06	0.03	0.678	0.322
SPART	0.12	0.04	**0.70**	-0.04	0.571	0.429
SPARTD	-0.02	-0.02	**1.01**	0.01	0.995	0.005
SDMT	0.24	0.37	0.26	0.05	0.450	0.550
PASAT	-0.05	0.01	-0.11	0.16	0.036	0.964
WLGT	0.08	**0.41**	0.24	0.04	0.348	0.652
VOSP2	-0.08	**0.82**	0.07	0.02	0.675	0.325
VOSP3	0.04	**0.50**	-0.13	0.01	0.240	0.760
BORB2	0.00	-0.06	-0.02	**0.82**	0.658	0.342
BORB3	0.08	0.29	-0.07	**0.42**	0.331	0.669
BORB4	-0.06	0.21	0.22	0.19	0.157	0.843
BORB5	0.10	0.15	0.10	**0.44**	0.314	0.686
GNT	0.10	**0.55**	-0.10	-0.03	0.315	0.685
KDT	0.19	0.24	0.00	-0.11	0.122	0.878
PPT	0.16	**0.49**	-0.14	-0.08	0.273	0.727
Proportion variance	15 %	12 %	10 %	7 %		
Cognitive functions assessed	Verbal learning and memory	Object naming, semantic storage	Visuo-spatial learning and memory	Visuo-perceptive functions		

Note. Loadings of 0.40 or higher are highlighted in each factor

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test)

Model-level explanatory power vs. parsimony: Given the exploratory nature of the analysis, four-, three- and two-factor solutions were evaluated. The Bayesian information criterion (BIC) was -243, -276.59, and -269.35 respectively. Although this supported the three-factor model as an optimum balance between explanatory value and parsimony, the three-factor model explained only 38% of

variance whereas the four-factor model explained 44% of variance. Since coverage of a broad range of cognitive functions was central to the overall project's success, the four-factor solution was preferred; and the factor loadings from the four-factor solution can be seen in Table 3.5.

d) Interpretation of the identified dimensions (factors)

The four-factor solution presented in Table 3.5 accounted for 44% of total variance. The most important factor (explaining 15% of the total variance) was related to verbal learning and memory, and consisted of the BRBN SRT task variables. The second factor (explaining 12% of the total variance) defined semantic object knowledge and consisted of the naming tasks. The third factor (10% total variance) was visuospatial learning and memory and was comprised of the SPART items. The fourth factor (7% total variance) comprised of items measuring pre-categorical properties of objects. This analysis was restricted to patients, however when repeated with the controls they resulted in no substantive changes for key findings.

III. A theory-based account of dimensionality in the phase I dataset

Classical psychology (encompassing cognitive neuropsychology) posits a modular structure of cognitive abilities (Fodor, 1983). The coverage of the phase I battery with respect to conventionally understood cognitive functions (dimensions) was presented in Table 3.6. The phase I battery contained tests of memory domain, sustained attention and concentration and processing speed domain, visuoperceptual and language domains. It also included two tests of audio-visual integration that theoretically would fall into both visuoperceptual and language domains as they involved visual processing and semantic knowledge about objects.

During the process of optimisation it was attempted to ensure that the phase II battery would retain the dimensionality and coverage of the phase I battery. This would therefore allow for valid comparisons between baseline and follow-up assessments to be made, and ensure that the longitudinal study investigated the cognitive changes separately in each cognitive domain.

Table 3.7. Theoretical account of the coverage of the phase I battery

Test item	Function	Domain
SRTL	Verbal learning, long-term storage	Verbal memory
SRTC	Verbal learning, consistent long-term retrieval	
SRTD	Verbal memory, delayed recall	
SPART	Visuospatial learning, long term storage	Visuospatial memory
SPARTD	Visuospatial memory, delayed recall	
SDMT	Visual scanning and tracking	Processing speed, attention and concentration, semantic fluency
PASAT	Auditory addition and inhibition	
WLGT	Spontaneous production of names in a category	
VOSP5	Space perception, single point localization	Visuoperceptual
VOSP6	Space perception, position discrimination	
VOSP7	Space perception, position identification	
VOSP8	Space perception, complex spatial relationships	
BORB2	Apperceptive recognition of length	
BORB3	Apperceptive recognition of size	
BORB4	Apperceptive recognition of orientation	
BORB5	Apperceptive recognition of position	
VOSP2	Object perception, Silhouette recognition and naming	
VOSP3	Object perception, object identification	
Orchard test	Audio-visual non-semantic processing	Audio-visual integration
SPMT	Audio-visual semantic processing	
GNT	Expressive semantics, object naming	Language
BC.Index	Expressive syntax, spontaneous language production	
PPT	Understanding semantic relationships between nouns	
KDT	Understanding semantic relationships between verbs	
TTT	Understanding semantic relationships between sequential events	
TROG	Reception of grammatical constructs	
MPNWDS	Receptive phonology, auditory non-semantic discrimination	

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

3.5. Optimization of the phase II cognitive assessment battery: a synthesis

In order to combine the preceding considerations and resolve a final phase II assessment battery, the following were prioritised (in order of importance): domain coverage (theoretical), domain coverage (data-driven), non-redundancy, detection of impairment, detection of variability, and practical considerations.

3.5.1. Domain coverage

The first PCA factor clustered together tasks that measure verbal learning and recall, and it consisted of three items and all of them were derived from the same SRT task. Since in order to collect the SRTL, SRTC and SRTD items it takes the same one administration procedure and they are all individually informative as they measure different aspects of verbal memory, it would be counter-productive to remove any of them, as the battery wouldn't become shorter to administer, but it would most definitely lose coverage. Therefore it was decided not to remove any of the first factor items.

The second factor clustered together the neuropsychological tasks that measure picture naming ability and storage of semantic knowledge about objects, but not the semantic knowledge about actions, as KDT did not load highly on this factor. Even though all of the items in factor 2, including WLGT, VOSP2, VOSP3, GNT and PPT, are administered slightly differently, as in Animal fluency the participants are required to name animals from memory, in GNT they need to name pictures of objects, in VOSP items the participants name shadows of objects, and in PPT participants match two objects by their meaning; however, all of these tests are believed to measure the same semantic storage function. Therefore it is reasonable to conclude that these items could be considered redundant and that some of them could be removed in order to abbreviate the battery.

The third factor clustered items from one test only, SPART, and included an item measuring the visuospatial learning, and an item measuring the delayed visuospatial recall. Similarly as with the items in factor 1, both of the items in factor 2 came from the same test and complement each other by measuring different components of visuospatial memory, and dropping one of them would have

only decrease the coverage of the battery without shortening its administration time. Therefore it was decided not to remove any of the third factor items.

The fourth factor clustered three items from the BORB battery: BORB2, BORB3 and BORB5. Contrastingly to the items from the factor 2, the factor 4 item properties were not tied to stored knowledge about the particular objects involved. In this instance they were directed solely at the perception of basic properties and object forms – those include their size, orientation, location and length. Since all of these items measure similar functions, they cannot be said to be entirely overlapping, as matching for length involves not entirely the same neuropsychological functions as matching for orientation and location, therefore these items could not be argued to be redundant. Therefore it was decided not to remove any of the items from factor 4.

From the PCA investigation of domain coverage I identified items that did not load on any of the factors, (*i.e.* all of the loadings for those items were below 0.35). Such items included SDMT, PASAT, BORB4 and KDT. Since SDMT was the only item assessing information processing speed, and the PASAT was the only item that assessed attention and concentration, it was decided to leave them in the battery. Furthermore, in the prevalence analysis SDMT proved to be a suitable test for picking up MS-related cognitive impairment, thus it would be unreasonable to exclude it based on the analysis of coverage only. However, the situation with the PASAT was not as clear – a moderate number of patients have failed it, but at the same time, it was too difficult for two of the controls as well. Since the PASAT is part of the BRBN battery and one of the most popular cognitive tests in MS research, it was decided to retain it for the phase II battery in order to be able to compare our results to other research studies. However, the same did not apply to the BORB4 and KDT tests. These tests have not been validated in MS and haven't been used in other studies, thus I couldn't even use them to compare our results to previous research. Therefore it was decided to exclude those items from the phase II battery.

3.5.2. Non-redundancy

From the analysis of redundancy I found that the factor 2 items WLGT, VOSP2, VOSP3, GNT and PPT measured the same function. I used the prevalence of impairment analysis (presented earlier in this chapter, section 3.4.2.) to decide which of those items to drop. As the least patients have failed the GNT and PPT items, those tests have been decided to be less effective measures of the semantic knowledge and therefore were decided to be removed.

3.5.3. Detection of impairment and variability

Items that have true frequencies of failure for MS patients higher that 10% were regarded as important to retain. However, of those with a true frequency of failure below 10% (in decreasing order: BORB3, PPT, BORB5, VOSP8, VOSP5, GNT, Orchard test and KDT), I elected to include BORB5 and VOSP8 to demonstrate contrast, pick up the new emerging impairments, and ensure sufficient coverage of the visuoperceptual domain.

3.5.4. Practical considerations

It was decided to remove the Orchard task and the SPMT task based on practical grounds as they simultaneously measure both auditory and visual functions and are therefore challenging to interpret. MPNWDS, KDT, TTT and PPT tests were omitted as they were considered to be experimental in their nature and had not yet been validated in MS samples. So far to the best of my knowledge they have only been used with MND patients and on small numbers. In addition, the KDT, PPT, TTT, Orchard, SPMT, and MPNWDS tasks required two separate laptops to administer and since data collection included travelling to patient homes carrying all of the materials with the researcher, for practicality reasons it was decided to avoid including these items.

The NART test was omitted from the phase II battery, as it was considered to be a measure of premorbid intelligence that (theoretically) should not change at re-assessment. Repeated administration of NART at phase II was considered redundant.

3.6. Inclusion of additional assessments not performed in phase I

Consideration was given to advances in the understanding of cognitive impairment in MS that have emerged in the time interval since design of the phase I study. Two principal areas were identified: cognitive reserve and self-perception of change.

3.6.1. Cognitive reserve

Recent publications in cognitive epidemiology increasingly stress the role that cognitive reserve plays in susceptibility to cognitive impairment and in the rate of progression of cognitive symptoms in both healthy and clinical populations. It was therefore decided to include a short assessment of MS-specific cognitive reserve and cognitive leisure activities questionnaire developed by Sumowski and colleagues (Sumowski et al., 2010).

3.6.2. Other advances of relevance

I also added a few questions that relate to factors that in the previous literature have been shown to have a positive effect on preventing cognitive decline. The first item related to bilingualism research, and I added a question that asks participants whether they speak any other languages besides English. The level of foreign language ability required to count the language as spoken is being able to ask for directions in the street and order a meal in the restaurant. The self-reported number of languages spoken was chosen to be recorded.

The second item added was related to physical activities and relates to previous research on benefits of exercising on concentration. In this study we asked the participants to indicate whether they take part in physical activities. Walking was not considered as exercise. More strenuous physical activities, such as running, swimming, cycling or hiking were counted as exercise, as well as such activities as Yoga, Pilates, stretching or muscle toning.

In the phase II study I've also chosen to collect data on participants' employment status. All participants were asked to indicate if they were employed at the point of phase II assessment, and

whether they had been employed at the point of phase I assessment. No distinction was made between part-time and full-time employment, and volunteering was not regarded as employment

3.6.3. Self-perception of changes in cognitive performance

Another type of evaluation added to the phase II battery was an assessment of the perceived progression of cognitive deficits. Performing this assessment allowed me to evaluate the cognitive change not only objectively, but also subjectively from the patient perspective.

The self-evaluation of performance at phase II was measured by asking the participants to estimate their test performance on a 100-point scale from 'very poor' to 'very good' after completing the test (e.g. how well you think you performed on these tests?). Then the participants were asked to indicate on the same scale from 0 to 100 how they think they had performed on the same tests at the baseline (phase I) assessment. If the participant indicated that they estimate they've performed around 50th centile at the follow-up and then that they think that they've performed at 70th centile at the baseline, this would indicate that the subject perceived a 20% decrease in one's cognitive ability in the time interval between the phase I and phase II assessments. The participants were informed that differences above 10% will be regarded as perceived change in their performance, so if they were unsure about whether they have changed or not they were asked to rate their performance at phase II in the range of -10 to +10 from their performance at phase I.

Having included the additional assessments of cognitive reserve and insight, I devised the final version of the optimised phase II battery. The items included into the phase II battery assessed physical ability; memory, processing speed, attention, semantic fluency; visuoperceptual abilities; receptive and expressive language abilities; self-perception of deficit progression, and questionnaires on cognitive reserve, depression, and MS impact on life, as well as additional questions about bilingualism and exercise.

3.7. Phase II assessment schedule

All phase II assessments were administered by the author at a single visit. Since the assessments of phase I and phase II were administered by two different researchers as two separate PhD projects, several steps were taken to ensure inter-observer reliability and comparability of the results. Those steps included administering the assessments in the same sequence and by using the same test-administration procedure and strictly following the test administration protocols developed by the test authors.

Table 3.8. Items included into the phase II assessment battery

Self-administered questionnaires	Cognitive tests	Clinical measures
Beck's Depression Inventory II (BDI-II)	Addenbrooke's Cognitive Examination – Revised (ACE-R)	Snellen chart
Multiple Sclerosis Impact Scale (MSIS-29)	Brief Repeatable Battery of Neuropsychological tests (BRB-N)	Kurtzke's Expanded Disability Status Scale (EDSS)
7-question cognitive reserve and cognitive leisure activities questionnaire (based on Sumowski et al., 2010)	Visual Object and Space Perception Battery (VOSP, only subtests 2, 3, 6 and 8)	
Questions on bilingualism and exercise	Birmingham Object Recognition Batter (BORB, only subtests 2 and 5)	
Self-Awareness of cognitive deficit scales	Test of Reception of Grammar (TROG)	
	Boston Cookie Theft Test (BCT)	

The phase II battery included the items presented in Table 3.7 and they were to be administered in the following order. First the demographic information and medical history were collected, followed by physical examination including assessment of visual acuity (Snellen chart) and neurological functions (EDSS). Then the ACE-R was administered, noting the participants with severe global cognitive impairment. After ACE-R the remaining neuropsychological tests were administered.

The first block of tests evaluated verbal and visuospatial memory, attention and concentration, information processing speed, semantic fluency, and was comprised of the BRBN items. The BRBN items were administered in the same order as in the baseline assessment: SRT learning, SPART

learning, SDMT, PASAT, SRT delay and SPART delay. The WLGT task again was used from the ACE-R battery. After having completed the BRBN the participants were asked to evaluate their performance on the BRBN tasks at phase II and at phase I retrospectively.

The second block of tests included the visuoperceptual tasks which were then administered in the following order: VOSP2, VOSP3, VOSP6, VOSP8, BORB2 and BORB5. Having done that the participants were again asked to give estimates of their performance on the visuoperceptual tests at phase II and at phase I.

The third block included the language tasks, administering the TROG first, followed by the BCT test. Having completed these two tasks, the participants were asked to provide estimates of their performance on the language tests at phase II and at phase I. This concluded the neuropsychological part of the evaluation.

Lastly the participants were asked to fill out the self-administered questionnaires: cognitive reserve questionnaire (with additional questions on bilingualism, exercise and employment), MSIS-29, and BDI-II, with the experimenter still present in the room. Once the questionnaires were completed, participants were debriefed and the phase II assessment ended.

3.8. Chapter summary

Chapter Three presents a description of the procedure behind the phase II protocol development. The rationale for optimizing the phase II battery had been identified through feedback from phase I participants who have found the phase I study to be too long and an unchanged assessment schedule was likely to impact negatively to phase II recruitment. Moreover, as advances in the understanding of cognitive impairment in MS have emerged since the phase I study was designed it was aimed to incorporate them into the phase II study schedule.

The methodology behind the phase II cognitive battery optimisation was based on empirical and theoretical justifications. The empirical aspect of the optimisation procedure included removal of redundant or insensitive tests while maintaining the domain coverage unchanged. Firstly this has been achieved through an investigation of the test item responsiveness and sensitivity to detect

variability attributable to MS. Secondly, an empirical evaluation of the dimensionality of the phase I battery was done through the principal component analysis; and through the analysis of the four identified domains we identified the test items that ensured adequate coverage while minimizing redundancy of the phase II battery. The theoretical aspect of the optimization procedure included a synthesis of these empirical findings against the descriptions of the tests with respect to conventionally understood cognitive functions. This was performed in order to ensure that while incorporating the empirical findings the phase II battery would retain the dimensionality and coverage of the phase I battery, and therefore allow for valid comparisons between baseline and follow-up assessments to be made. Moreover, in that section the additional assessments were also incorporated into the phase II battery responding to advances in the field that have occurred since phase I, namely the items assessing cognitive reserve, protective factors against cognitive decline, and self-awareness measures of changes in the cognitive performance.

Throughout the battery optimization procedure it was achieved that the changes made to the study battery did not affect any other test parameters, such as the test setting and sequence, and all attempts were undertaken to ensure that the results from the phase I and phase II would be comparable.

The optimized phase II battery was considered suitable for follow-up investigations as it was reduced to three hours (making it manageable to be administered during a single visit) and it had good coverage of cognitive functions, and the clinical and non-clinical variables associated with cognitive performance.

Chapter Four. Characterization of the phase II cohort

4.1. Chapter overview

This chapter describes implementation of the phase II protocol, including characterisation of the cohort who consented to participation at phase II (follow up) and an evaluation of comparability with respect to the phase I cohort.

4.2. Implementation of the phase II study protocol

4.2.1. Approvals

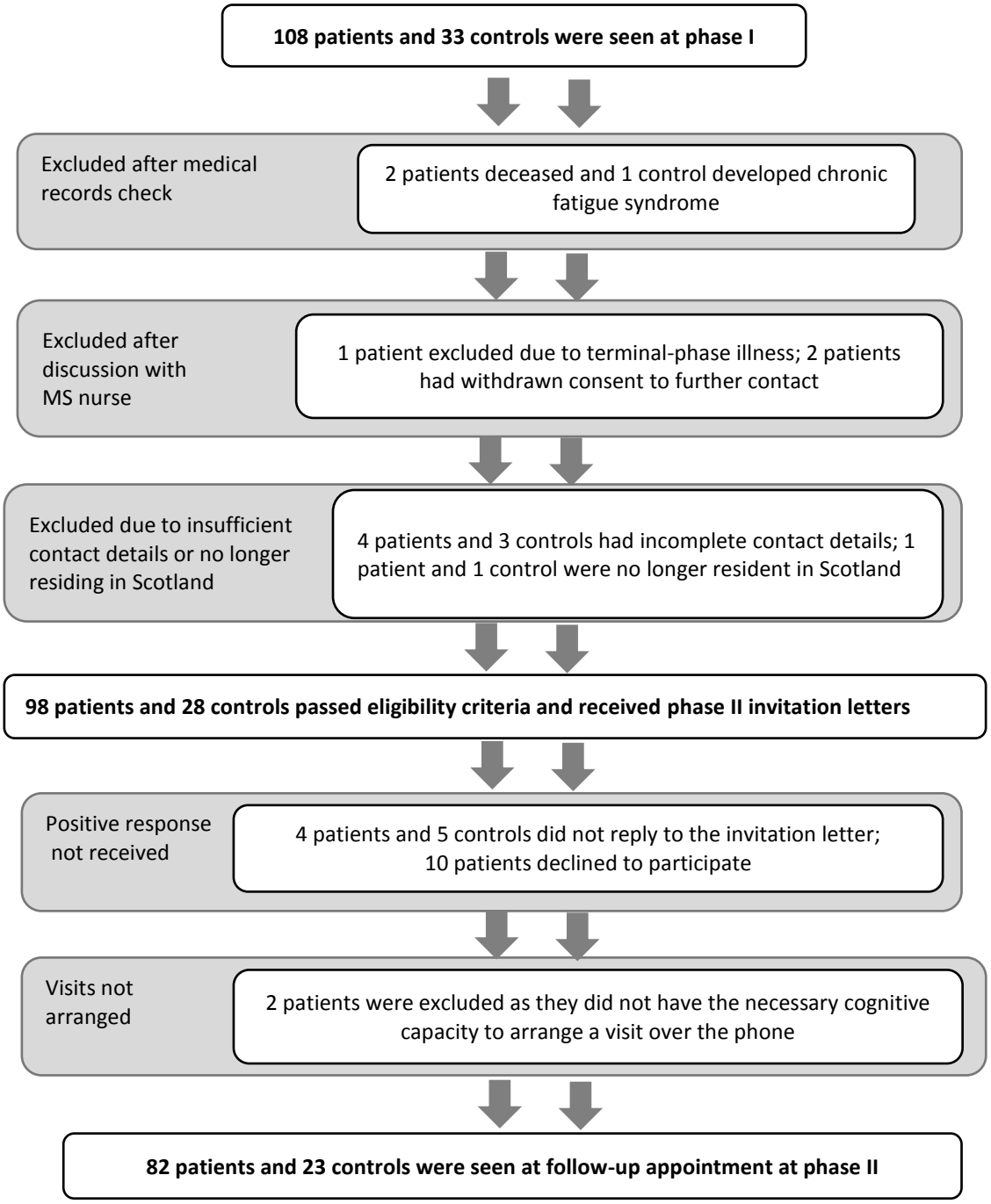
Approval for the phase II study was requested as minor amendment to the (phase I) Phenotypic Characterisation of Cognition in Multiple Sclerosis (MS) study (REC Reference number 10/S1103/54, NHS R&D reference number: 2010/W/NEU/14) from the South East Scotland Research Ethics Committee II. Ethical approval was received on 3rd July, 2014. Approval from NHS R&D was received on 18th August, 2014.

4.2.2. Case ascertainment & recruitment

The screening process was applied as described previously (Chapter Three, section 3.2.6). Of the 108 pwMS and 33 controls in phase I, 82 pwMS and 23 controls were recruited to phase II (Figure 4.1). The raw recruitment rate was therefore 76% for patients and 70% for controls. Of those who fit the recruitment criteria and were invited to participate in phase II (98 pwMS & 28 controls), recruitment rates were 84% and 82% respectively.

Notably, two phase I participants were excluded from phase II as they did not have the sufficient cognitive capacity to give consent to participate. That was considered to be evidence that those individuals had progressed cognitively even if their details were not collected at phase II. The issue of phase II sample representativeness with regards to the phase II non-participants will be addressed later in the discussion chapter (Chapter Eight)

Figure 4.1. Flowchart of participant journey from phase I to phase II



4.2.3. Systematic biases in recruitment to phase II

Given that phase II participants were selected through a combination of pre-specified eligibility criteria and self-selection (informed consent), potential systematic (non-random) differences may have arisen between those who participated in phase II and those who did not.

Comparison of these groups of pwMS is shown in Table 4.1. Those patients who did not participate in phase II were more likely to be female (X^2= 5.356, p = 0.021) and exhibit higher levels of self-reported depressive symptomatology (U = 670.00, p = 0.019) at baseline.

Table 4.1. A comparison between demographic and disease variables at baseline for pwMS who were phase II participants and non-participants

	Phase II participants (n=82)	Phase II non-participants (n=26)	Difference
Age	48.38 (9.52)	46.00 (10.35)	U=924.50, p=0.309
Gender (% female)	**52%**	**69%**	**X^2=5.356, p=0.021**
Years of education	12.83 (3.06)	11.96 (2.37)	U=807.50, p=0.305
Dementia (ACE-R <82)	6% yes, 94% no	12%yes, 88% no	X^2 = 0.852, p=0.356
BDI-II depression score	**14.62 (9.17)**	**21.68 (13.07)**	**U=670.00, p=0.019**
MS disease course	24% PP, 34% SP, 46% RR	8% PP, 35% SP, 57% RR	X^2=3.057, p=0.217
Disease duration	10.90 (8.08)	10.46 (7.53)	U=1046.50, p=0.888
EDSS score	4.53 (2.14)	4.73 (2.31)	U=997.50, p=0.667
MS modifying treatment	22%	38%	X^2=2.572, p=0.109
Antidepressant use	40%	58%	X^2 = 2.506, p=0.113

Table 4.1 shows that women and those with higher levels of self-reported depressive symptomatology were less likely to participate in phase II. Comparisons were done using the Mann-Whitney U test for interval and ordinal data, and Pearson's chi square test for frequency data.

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI-II – Beck's Depression Inventory 2nd Ed., EDSS – Expanded Disability Status Scale

Comparison of the groups of control individuals who participated in phase II assessments and those who did not is shown in Table 4.2. No statistically significant differences were found in age, gender, years of education, dementia or depressive symptomology between the controls who participated in the phase II and those who didn't.

Table 4.2. A comparison between demographic variables at baseline for controls who were phase II participants and non-participants

	Phase II participants (n=23)	Phase II non-participants (n=10)	Difference
Age	48.87 (12.25)	41.70 (13.27)	U=78.00, p=0.147
Gender (% female)	65%	70%	X^2=0.072, p=0.789
Years of education	12.43 (2.33)	12.60 (2.41)	U=112.50, p=0.920
Dementia (ACE-R <82)	0	0	
BDI-II depression score	3.27 (2.79)	4.50 (3.69)	U=16.00, p=0.427

Table 4.2 shows no statistically significant difference in the demographic characteristics of those who participated and did not participate in phase II. Comparisons were done using the Mann-Whitney U test for interval and ordinal data, and Pearson's chi square test for frequency data

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI-II – Beck's Depression Inventory 2nd Ed.

4.2.4. The phase II study environment

The phase II assessment environment was considered to be comparable at all research locations (Anne Rowling Clinic, Psychology Department research laboratories and patient homes). In the Anne Rowling Clinic the study participants were seen in a quiet and well-lit room with a desk; a research nurse was present on-site during the assessments. At home, subjects were seen seated in their kitchens or living rooms at a desk; all potential distractors, such as pets, TV or family members, were removed, and lightening was adjusted. The conditions in the research laboratories were no different from the conditions at the Anne Rowling Clinic, as these rooms were quiet, well-lit and had no distractions.

Two of the controls (partner and daughter of patients) were seen at home, and the remaining twenty-one were seen at the research laboratories of The University of Edinburgh Psychology Department.

Thirty-one patients were seen at the Anne Rowling Clinic and the remaining fifty-one were seen at home. As it can be seen from Table 4.3, the patients who preferred to be seen at home had more neurological disability (higher EDSS scores; U = 545.50, p=0.014) and longer MS duration (U = 528.50, p = 0.020). The frequencies of MS courses was uneven (X^2 = 10.089, p = 0.006), with more patients with SPMS preferring to be seen at home. In addition, the patients seen at home had much

more cognitive difficulties as 22% of them performed below the ACE-R cut-off for dementia (X^2 = 7.722, p = 0.005) and had fewer years of education (U = 510.50, p = 0.030) than those seen in the clinic.

Table 4.3. A comparison between demographic and disease variables at phase II for patients who were assessed at the clinic *vs.* those assessed at their home

	Seen at the clinic (n=31)	**Seen at home (n=51)**	**Difference**
Age	50.22 (8.11)	51.82 (10.31)	U=718.50, p=0.491
Gender (% female)	42%	59%	X^2=2.205, p=0.138
Years of education	**13.87 (3.46)**	**12.21 (2.65)**	**U=510.50, p=0.030**
Dementia (ACE-R <82)	**0**	**22%**	**X^2=7.722, p=0.005**
BDI-II depression score	13.74 (8.98)	16.22 (9.69)	U=640.50, p=0.191
MSIS-29 score	80.55 (27.16)	85.98 (26.29)	U=691.00, p=0.414
MS disease course	**32% PP, 13%SP, 55% RR**	**18% PP, 47% SP, 35% RR**	**X^2=10.089, p=0.006**
Disease duration	**10.96 (4.58)**	**15.92 (9.13)**	**U=528.50, p=0.020**
EDSS	**4.58 (2.04)**	**5.82 (1.79)**	**U=545.50, p=0.014**
MS modifying treatment	58%	62%	X^2=0.124, p=0.725
Antidepressant use	35%	40%	X^2=0.134, p=0.714

Table 4.3 shows that phase II participants with MS who were seen at home were less well-educated, had a higher rate of dementia-level cognitive impairment on the ACE-R, longer duration of disease, higher levels of physical disability, and were more likely to be in the secondary progressive phase than relapsing-remitting phase. Comparisons were done using the Mann-Whitney U test for interval and ordinal data, and Pearson's chi square test for frequency data

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI-II – Beck's Depression Inventory 2nd Ed., MSIS-29 – 29 item Multiple Sclerosis Impact Scale, EDSS – Expanded Disability Status Scale

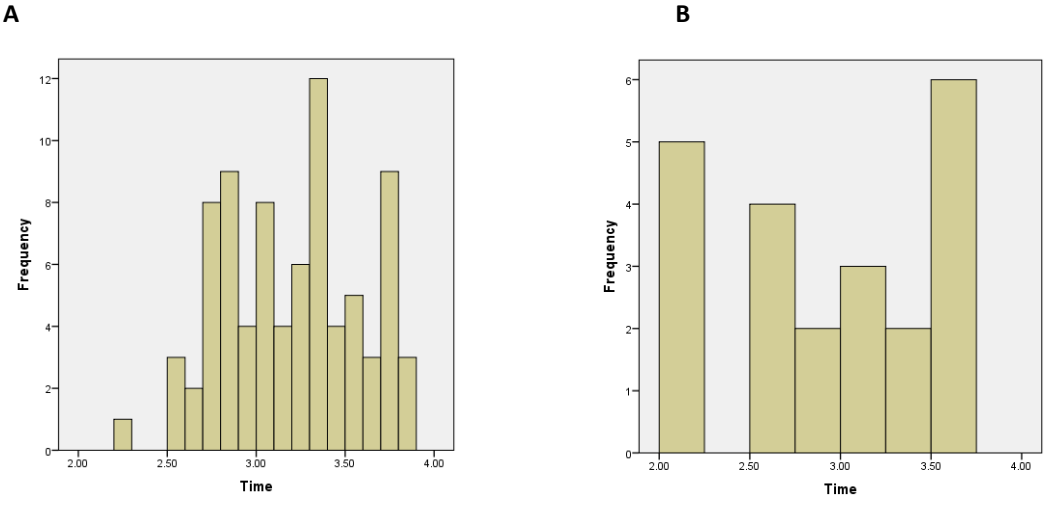
4.2.5. Tolerability of phase II assessment

82 patients and 23 controls have completed the phase II assessments. Two patients did not manage to go through the phase II assessments in one appointment, and they were seen again to finish with the assessment one week later. Three patients could not take part in tasks that required upper limb movements due to severe neurological disability. One participant had moderate diplopia and could not perform some of the visuoperceptual tasks. Two patients had their cognitive deficits progressed to the level where they did not understand the requirements of some of the more difficult tasks, and those tasks were therefore not administered. All other tasks were performed to completion.

4.2.6. Time interval in-between phase I and phase II assessments

82 patients and 23 controls completed both phase I and phase II assessments. The distribution of time intervals in-between assessments is shown in Figure 4.2. The mean time between phase I and phase II assessments for the patients was three years and two months (SD ± 4.5 months) and for the controls the mean time was two years and eleven months (SD + - 6.7 months).

Figure 4.2. Time interval in years in-between phase I and phase II assessments. The data for patient (A) and control (B) participants is presented



The time duration in-between assessments was shorter for the control sample (t[26.42] = 2.193, p = 0.037). This difference will be considered when comparing longitudinal changes in the patient sample with the control sample.

4.3. Characterisation of the phase II dataset & cohort

4.3.1 Data completeness

The phase II assessment data had high completeness with very few missing observations (Table 4.4). The completeness for the patient sample was 98.15% and for the control sample it was 100%.

All the missing observations reflected patients with specific disabilities. The patients with upper limb paralysis were unable to perform the SPART test as the checkers needed to be moved. The VOSP6 and BORB5 items were not administered to the patient with severe diplopia. 14 patients had found the PASAT rules too difficult to understand due to advanced cognitive impairment; they did not pass the practice, therefore this test was not performed on them. One patient was functioning so badly that she was able to understand only a small number of task requirements and couldn't go beyond performing a few of the ACE-R tasks. She had profound physical impairments of all limb functions and severe head tremor (EDSS 9.5). Therefore all cases of missing data can be explained and thus argued not to be missing at random.

Table 4.4. Phase II data completeness

Variable	Patient data completeness (% of 82)	Control data completeness (% of 23)
Demographical variables:		
Age	100%	100%
Gender	100%	100%
Years of education	95%	100%
Covariates of interest:		
Pre-morbid cognitive leisure activity score	99%	100%
Employment	100%	100%
Languages	99%	100%
Exercise	100%	100%
Disease variables:		
Disease duration	100%	NA
EDSS score	100%	NA
MS type	100%	NA
MSIS-29 score	99%	NA
BDI-II depression score	99%	100%
Use of MS modifying treatment	99%	NA
Use of antidepressants	96%	NA
Cognitive test scores:		
ACE-R	100%	100%
SRTL	99%	100%
SRTC	99%	100%
SRTD	99%	100%
SPART	95%	100%
SPARTD	95%	100%
SDMT	99%	100%
PASAT	83%	100%
WLGT	100%	100%
VOSP2	100%	100%
VOSP3	99%	100%
VOSP6	96%	100%
VOSP8	99%	100%
BORB2	99%	100%
BORB5	96%	100%
TROG	99%	100%
BC.Index	98%	100%
TOTAL DATASET	**98.15%**	**100%**

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI-II – Beck's Depression Inventory 2nd Ed., MSIS-29 – 29 item Multiple Sclerosis Impact Scale, EDSS – Expanded Disability Status Scale, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), NA – not applicable

4.3.2. Phase II participant characteristics

82 patients and 23 controls had participated in both phases of the follow-up study. Summaries of the patient and control cohorts at each phase are shown in Table 4.5 and Table 4.6 respectively.

Table 4.5. Demographic and disease characteristics of participants with MS from the follow-up sample at phases I & II

	Phase I (n=82)	Phase II (n=82)	Difference
Gender (% female)	43 (52%)	43 (52%)	p = 1.000
Years of education	12.83 (3.06)	12.85 (3.07)	Z = -1.000, p = 0.317
Employed (%)	**42 (51%)**	**27 (33%)**	**p < 0.001**
Dementia (ACE-R <82)	**5 (6%)**	**11 (13.4%)**	**p = 0.031**
BDI-II depression score	14.62 (9.16)	15.27 (9.45)	Z = -0.683, p = 0.495
MS disease course	19 (23%) PP, 38 (46%) RR, 25 (31%) SP	19 (23%) PP, 35 (43%) RR, 28 (34%) SP	X^2 = 3.00, p = 0.083
Disease duration	**10.90 (8.08)**	**14.09 (8.09)**	**Z = -8.275, p < 0.001**
EDSS	**4.53 (2.14)**	**5.35 (1.97)**	**Z = -4.006, p < 0.001**
MS modifying treatment	**18 (23%)**	**49 (62%)**	**p < 0.001**
Antidepressant use	27 (38%)	29 (41%)	p = 0.815

Table 4.5. shows that at phase II, patients were less likely to be in employment, but more likely to have a higher rate of dementia-level cognitive impairment on the ACE-R, and be on MS modifying treatment. Levels of physical disability (EDSS) were also higher at phase II
Note: Wilcoxon Signed Ranks for interval and ordinal data, and 2x2 McNemar test and 3x3 McNemar-Bowker's test for frequency data employed
Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI – II – Beck's Depression Inventory, 2nd Ed., EDSS – Expanded Disability Status Scale

Table 4.6. Demographic and other characteristics of control participants at phases I & II

	Phase I (n=23)	Phase II (n=23)	Difference
Gender (% female)	14 (61%)	14 (61%)	p = 1.000
Years of education	12.43 (2.33)	12.65 (2.35)	Z = -1.342, p = 0.180
Employed (%)	18 (78%)	17 (74%)	p = 1.000
Dementia (ACE-R <82)	0	0	p = 1.000
BDI-II depression score	3.27 (2.79)	3.69 (3.20)	Z = -1.929, p = 0.054

Table 4.6. shows that there had been no significant changes in the characteristics of control sample.
Note: Wilcoxon Signed Ranks for interval and ordinal data, and 2x2 McNemar for frequency data employed
Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI – II – Beck's Depression Inventory, 2nd Ed., EDSS – Expanded Disability Status Scale

As expected, in the time period in-between phase I and phase II assessments there had been a general increase in measures associated with progression of MS. Three patients have progressed from RRMS to SPMS and an overall increase in disability (EDSS score) was observed. Thirty-one more patients started using disease modifying treatments. Six more patients had developed moderate cognitive impairment as indicated by the performance below the cut-off sensitive to dementia in a dementia-screening tool (ACE-R < 82). The proportion of patients who were employed declined markedly between assessments from 51% at baseline to 33% at follow-up ($X^2(1) = 38.34$, $p < 0.001$) with all of the retirements being due to medical reasons. There were no other changes related to the course of MS in our patient sample. Besides MS, at phase II thirty-one patients had additional illnesses, such as high cholesterol, osteoporosis, arthritis, joint hypermobility, thyroid problems, depression, sleep apnoea and diabetes.

There had been no significant changes in the characteristics of the control sample.

4.3.3. Phase II patient- and control-group matching for demographic and other variables

The comparison of demographic and other characteristics in patient and control samples can be seen in Table 4.7.

Table 4.7. Demographic and other characteristics of control participants and patient participants at phase II

	Controls (n=23)	Patients (n=82)	Difference
Age	51.39 (11.87)	51.22 (9.52)	U=890.00, p=0.681
Gender (% female)	14 (61%)	43 (52%)	$X^2 = 0.514$, p= 0.473
Years of education	12.65 (2.35)	12.85 (3.07)	U=888.50, p= 0.945
Employed (%)	**17 (74%)**	**27 (33%)**	**$X^2 = 12.394$, p< 0.001**
Dementia (ACE-R <82)	0	11 (13.4%)	$X^2 = 3.493$, p = 0.062
BDI-II depression score	**3.69 (3.20)**	**15.27 (9.45)**	**U=181.00, p< 0.001**

Table 4.7. shows that pwMS in phase II were more likely to be unemployed and report higher levels of depressive symptomatology than controls. Note: Mann Whitney U-test for interval and ordinal data, and Chi Square test for frequency data were employed.

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI – II – Beck's Depression Inventory, 2nd Ed.

The phase II patient and control samples were well matched on demographic variables (age, gender, and years of education). The employment rate of control participants (74%) was more than two times higher than the patients' (33%) ($X^2 = 12.394$, $p < 0.001$), and the control employment rate at phase II was comparable to Scotland's national employment rate (73.5%) (OECD, 2014). The patients were more prone to moderate cognitive impairment (ACE-R<82) ($X^2 = 3.493$, $p = 0.062$) and had more depressive symptomology ($U=181.00$, $p< 0.001$) than the controls.

4.4. Chapter summary

This chapter described the implementation of the phase II protocol, including the characterization of the phase II cohort and an evaluation of their comparability with respect to the phase I cohort. Even though the participants were seen in the Clinic as well as at home, the phase II environment was considered to be comparable at all research locations. The phase II battery was shown to be highly tolerable and 98% of patients and 100% of controls have managed to go through it in a single visit. The average time between phase I and phase II assessments was approximately three years.

During the phase II of this study we managed to collect follow-up data on 82 pwMS and 23 controls ensuring high (over 70%) follow-up recruitment rates, and very high data completeness (over 98% for the patients and 100% for controls) at phase II.

The patient and control samples were well-matched in terms of demographic and health variables, although not surprisingly, the patients had been more depressed and had higher chances of being unemployed. There were no major differences between the phase II control participants and non-participants. However the phase II patient non-participants had a tendency to have slightly higher levels of depression, and more likely to be female. We have also compared the patients who had their phase II assessments performed in the clinic to those who had preferred to be seen at home. The patients who were seen at home had fewer years of education, were more neurologically and cognitively impaired, more often had a SPMS course and more years of MS duration.

In the three years in-between assessments a tendency was observed for the patients to develop neurological and cognitive impairments, lose employment, and start MS modifying treatment. The controls, however, remained relatively unchanged. Therefore it can be concluded that the changes

observed between the phase I and phase II performance in-between the two groups could be attributable to the variables associated with MS.

Chapter Five. The Dimensionality of Cognitive Impairment in MS

5.1. Chapter overview

This chapter addresses the issue of disentangling whether cognitive impairment in pwMS should be considered a unidimensional or multidimensional construct, and the longitudinal pattern of changes in domain-specificity of cognitive impairments.

This chapter starts with an attempt to investigate the domain-specificity of cognitive disturbances. The second but linked analysis revolved around identification of potential subgroups (clusters) of individuals (within the study population) who display distinct patterns of cognitive impairments and/or change. The dimensionality of the phase I and phase II datasets were analysed separately, then later the differences between phase I and phase II dimensionality were analysed as a third (longitudinal change) dataset. After each analysis I also explored clinical and demographic associations with the identified clusters of individuals.

5.2. Methods

The *dimensionality of cognitive impairments within the study population* was evaluated by examining both independence and overlap of domain-specific deficits. Cognitive domains (and the test items which evaluate them) were defined as:

- Verbal memory (SRTL, SRTC, SRTD)
- Visuospatial memory (SPART, SPARTD)
- Processing speed (SDMT, PASAT, WLGT)
- Visuoperceptual (VOSP 2, 3, 6 & 8; BORB 2 & 5)
- Language (TROG, BC.Index)

Even though the items in the processing speed domain (SDMT, PASAT, WLGT) measure a spectrum of cognitive functions: attention and concentration, executive functions and information processing speed; for practical purposes these three items were grouped together and analysed as one entity. It

was considered not to be a major mistake to group these tests under the name 'processing speed' as these were the only items in the phase I and phase II test batteries where the participants had to produce responses under the time pressure.

To evaluate the *dimensionality of the study population with respect to cognitive impairments*, a multivariate 'data driven' approach was used. Specifically, cluster analysis with cognitive tests as statistical units and subjects as variables listing the observations was employed.

5.2.1. Data format

In order to identify the cognitive dysfunction of potential clinical significance to pwMS, only discrete *cognitive impairments* (performance below a reference threshold – as discussed below) have been considered in this chapter. Quantitative *cognitive performance* (including normal range) is addressed in Chapter Six.

Cognitive performance data was coded ordinal binary into '0's and '1's. '0's represented performance above the cut-off score and '1's represented performance below the cut-off score for each subtest.

I. Defining normal cognitive performance

The lack of robust external normative data has already been noted (Appendix A). In order to determine the need for separate definitions (thresholds) of cognitive impairment at phases I and II, I therefore compared performance of 23 control subjects with both phase I and phase II data on the same cognitive tests using related measures non-parametric tests (Wilcoxon matched-pairs test) (Table 5.1). A non-parametric test was chosen as the controls tend to perform near ceiling on a number of tests with little variation, with item distribution being significantly negatively skewed.

On the majority (9/16 = 56%) of tests the controls showed no statistically significant difference in performance between phase I and phase II assessments. However, the controls have showed improvement on visuospatial learning, information processing speed and semantic naming items; and deterioration on spatial orientation and expressive syntax items.

Table 5.1. Analysis of difference in control performance on the same tests at phase I and phase II of the longitudinal assessment.

	Phase I (n=23)	Phase II (n=23)	Difference
ACE-R	96.96 (2.72)	96.96 (1.82)	Z = -0.066, p = 0.947
SRTL	49.78 (12.10)	49.78 (11.42)	Z =-0.293, p = 0.770
SRTC	39.96 (14.25)	40.09 (13.48)	Z = -0.228, p = 0.819
SRTD	9.91 (2.13)	9.26 (2.32)	Z = -1.256, p = 0.127
SPART	**20.94 (4.95)**	**22.91 (4.21)**	**Z = -2.198, p = 0.028**
SPARTD	7.30 (1.82)	8.04 (2.46)	Z = -1.576, p = 0.115
SDMT	**54.32 (6.93)**	**57.61 (7.26)**	**Z = -2.808, p = 0.005**
PASAT	47.59 (12.70)	50.70 (6.55)	Z = -1.114, p = 0.265
WLGT	**22.00 (4.28)**	**24.04 (5.10)**	**Z = -2.096, p = 0.036**
VOSP2	**23.22 (3.20)**	**24.91 (3.30)**	**Z = -2.630, p = 0.009**
VOSP3	18.70 (1.26)	18.48 (1.47)	Z = -0.408, p = 0.683
VOSP6	**19.83 (0.49)**	**18.30 (1.18)**	**Z = -3.668, p < 0.001**
VOSP8	9.91 (0.28)	9.65 (0.57)	Z = -1.897, p = 0.058
BORB2	27.39 (1.30)	27.35 (1.85)	Z = -0.029, p = 0.977
BORB5	**37.13 (2.34)**	**36.22 (2.23)**	**Z = -2.070, p = 0.038**
TROG	38.81 (1.50)	38.52 (1.47)	Z = -1.493, p = 0.135
BC.Index	**41.68 (4.57)**	**34.06 (7.56)**	**Z = -3.847, p < 0.001**

Table 5.1. shows that the controls have showed improvement on visuospatial learning, information processing speed and semantic naming items; and deterioration on spatial orientation and expressive syntax items.

Note. All comparisons were performed using the Wilcoxon matched-pairs test

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD– Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

The substantial heterogeneity of normative (control) performance between phase I and II confirmed the need to define separate normal ranges for all tests at phases I and II so that valid attribution of changes in the MS population could be made to disease rather than experimental variation, practice, or aging effects.

II. Thresholds for impairment at phase I

As previously described (Chapter Three, Table 3.3.), the definition of cognitive impairment at phase I was based on the performance of 29 controls, and was further defined as performance below 2 SD below phase I control mean.

III. Thresholds for impairment at phase II

Two SD thresholds were also applied to the phase II control cohort (n=23), and gave differing values for eleven (of sixteen) tests (Table 5.2).

Table 5.2. Cut-off values used for definition of the range of normative performance for phases I and II. These values were calculated based on 2SD below phase I and phase II control means

Test	Phase I (n=29)	Phase II (n=23)
SRTL	28	27
SRTC	13	14
SRTD	6	5
SDMT	40	44
PASAT	18	38
SPART	12	15
SPARTD	4	4
WLGT	15	14
VOSP2	17	19
VOSP3	16	16
VOSP6	20	16
VOSP8	10	9
BORB2	24	24
BORB5	32	32
TROG	36	36
BC.Index	32 %	19 %

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD– Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

Six (6/16 = 38%) tests (SRTL, SRD, WLGT, VOSP6, VOSP8, BC.Index) had a lower threshold for impairment – *i.e.* the controls deteriorated or the test administration procedure made them more challenging. Five (5/16 = 31%) tests (SRTC, SDMT, PASAT, SPART, VOSP2) had a higher threshold to define impairment – *i.e.* the controls improved (practice effect?) or the test administration procedure have made them simpler.

5.2.2. The method of cluster analysis

All cluster analyses were performed using Stata/SE 9.2 for Macintosh (Stata Corp, TX, USA). Guidelines by Aldenderfer and Blashfield (Aldenderfer & Blashfield, 1984) for performing cluster analysis were followed. The Jaccard similarity measure was chosen as it has been shown to be suitable for working with binary data due to its ability to classify correctly not only the observations with the trait visible, but also the observations where the trait is not expressed[1]. This is particularly important in this study as not only we were interested in grouping patients with certain cognitive impairments together, but also we wanted to identify a subgroup of 'cognitively intact' patients.

Hierarchical clustering was chosen to use since it has been shown to be suitable for dichotomous data[2], smaller sample sizes, and to be less sensitive to outliers and non-normality[3]. In addition, hierarchical method has been recommended to be performed first in all analyses to determine the number of clusters (Milligan, 1980).

In this analysis the average linkage agglomeration method was chosen to employ instead of Ward's method, since the latter has been reported to struggle with clusters of different sizes (Kaufman & Rousseeuw, 2009). Since I expected language and visuoperceptual impairments to be less prevalent in our sample than information processing speed or memory impairments, it was presumed that the clusters would differ in their size.

---

[1] ("Similarity Coefficients for binary data: Why choose Jaccard over Russell and Rao?," 2013)

[2] ("Hierarchical or Two-step cluster analysis for binary data?," 2014)

[3] ("Comparing hierarchical clustering dendrograms obtained by different distances & methods," 2013)

The cluster analysis method tends to assume the absence of missing data and is only able to include the observations for which every variable was measured.

26 of 108 individuals with phase I data were removed from the following analyses as they did not have follow-up phase II data, as it was considered necessary to ensure that the same individuals were included into all analyses. Of the 82 patients with both phase I and phase II data, 21 had missing observations on at least one of the variables at either phase I or at phase II, and those individuals with incomplete data were therefore removed. Since the following analyses investigated the dimensionality of cognitive impairments, it was necessary to ensure that all participants had complete data in all assessments.

The dimensionality analyses were conducted on 61 individuals that had full data on the following items: SRTL, SRTC, SRTD, SPART, SPARTD, SDMT, PASAT, WLGT, VOSP2, VOSP3, VOSP6, VOSP8, BOR 2, BORB5, TROG and BC.Index.

However, in order to investigate the robustness of results from the complete case analysis, all analyses were repeated on the full dataset (n=82) after performing multiple imputations (Appendix E).

5.2.3. Number of clusters

In order to select the best suited number of clusters for our data it was chosen to employ the dendrite method. Since the optimal classification of objects occurs when the variances of between clusters to within clusters changes drastically between the levels of cluster analysis (Sneath & Sokal, 1973), we chose to utilize the F-statistic proposed by Caliński and Harabasz (Caliński & Harabasz, 1974); as it has been shown to be most fit for this task (Milligan & Cooper, 1985). The number of clusters is considered to provide more distinct clustering for the data when the Caliński and Harabasz F-statistic is high.

5.2.4. Identifying what the clusters represent

In order to determine what the clusters represent it was chosen to count how many of each type of cognitive assessments the individuals in each cluster have failed, and then based on that to make assumptions about the dimensionality of cognitive impairment for the individuals in each cluster.

5.2.5. Characteristics of individuals in each cluster

In this section for each analysis the demographic and clinical variables were explored in the attempts to explain the cluster groupings. The purpose of this section was purely explorative, hence no predictions were made.

5.3. The dimensionality of cognitive impairments

The result analysis of this chapter begins with an attempt to investigate the pattern in which the cognitive functions are affected by MS. First the dimensionality of cognitive impairments of phase I and phase II datasets is analysed separately, then later the dimensionality of emerging and resolving deficits as a third longitudinal dataset is studied.

5.3.1. Domain-specificity of impairments at phase I

The first step was to investigate the pattern of domain deficits at phase I. This analysis of dimensionality allowed to explore two elements – the cumulative burden of impairments (i.e. were they predominantly single-domain, or multi-domain), and the pattern of impairments (i.e. was there obvious over-representation of certain combinations of cognitive domain failure).

The definition of a threshold for domain impairment was arbitrary and represented a trade-off between sensitivity and specificity, especially since the domains had an unequal number of cognitive tests in them. An approach was chosen to define cognitive impairment on a domain when a patient failed at least two test items within a domain. The motivation behind this approach was that it was less sensitive but more specific than failing at least one item, while allowing comparable application across all tested domains. This approach was also considered to minimize the chances of picking up false positives, where the patients might have failed one test in a domain because they didn't

understand the task or due to fatigue. Being more conservative and avoiding such false positives was considered to be an important methodological aspect, since this study could be argued to be experimental in its nature as some of the tests employed here (e.g. majority of visuoperceptual and language tests) have never before been applied in MS research.

Table 5.3. The pattern of domain deficits for pwMS at phase I.

Cognitive domain	Number of items in each domain	Minimum number of tests to fail to be considered impaired on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

0 Domain unaffected. Represents performance below cut-off on one or less tests in that domain.

1 Domain impaired. Represents performance below cut-off on two or more tests in the domain.

Total domain impairments	Verbal memory	Visuospatial memory	Processing speed	Visuo-perceptual	Language	Number of pwMS
No impairments	0	0	0	0	0	33
Impairment on only one domain	1	0	0	0	0	8
	0	1	0	0	0	2
	0	0	1	0	0	1
	0	0	0	1	0	4
	0	0	0	0	1	
Impairment on two domains	1	1	0	0	0	
	1	0	1	0	0	4
	1	0	0	1	0	4
	1	0	0	0	1	1
	0	1	1	0	0	
	0	1	0	1	0	
	0	1	0	0	1	
	0	0	1	1	0	1
	0	0	1	0	1	
	0	0	0	1	1	1
Impairment on three domains	1	1	1	0	0	
	1	1	0	1	0	
	1	1	0	0	1	
	1	0	1	1	0	
	1	0	1	0	1	1
	1	0	0	1	1	
	0	1	1	1	0	
	0	1	1	0	1	
	0	1	0	1	1	
	0	0	1	1	1	
Impairment of four domains	1	1	1	1	0	1
	1	1	1	0	1	
	1	1	0	1	1	
	1	0	1	1	1	
	0	1	1	1	1	
Global impairment	1	1	1	1	1	

As it can be seen from Table 5.3, at phase I the patients tend to fail all of the cognitive domains, although the impairments on verbal memory items were the most common. This could be because the SRTL and SRTC variables have been highly inter-correlated, and failing one of them in most cases went together with failing another. Besides that there seemed to be no over-representation of a particular combination, suggesting that cognitive deficits at phase I could indeed be considered multidimensional. At phase I fifteen patients failed only one cognitive domain, eleven patients failed two cognitive domains, and two patients could be considered more globally impaired with one of them failing three, and another one failing four cognitive domains.

Another interesting aspect was that around half of patients at phase I could be considered cognitively spared, i.e. none of their impairments (if any) had qualified for failing a cognitive domain. However, it can be acknowledged that the impairment definition employed in these analyses was very conservative. In fact of those 33 patients, ten failed BC.Index, eight failed VOSP6, six failed SRTD, with smaller frequencies of them failing other tests. In addition, some of these patients in the 'no impairments' group had actually failed more than one item, although on different domains, thus still didn't qualify for a domain impairment. Two of those 'unimpaired' pwMS had failed four (of sixteen) tests; four failed three tests; five failed two tests, and eleven failed one test.

5.3.2. Domain-specificity of impairments at phase II

Now that I've investigated the phase I, the next step was to check whether the dimensionality of cognitive impairments has changed at phase II. Indeed, even with such conservative definition of domain-impairment, more patients have developed multi-dimensional deficits, as in phase II ten patients have failed three or more cognitive domains. Another interesting finding was that eight patients with impairments on three cognitive domains had identical patterns – they all have failed verbal memory, processing speed and visuoperceptual groups of tests. The one patient who had failed four cognitive domains has also had the same pattern with an additional impairment of visuospatial memory.

Table 5.4. The pattern of domain deficits for pwMS at phase II

Cognitive domain	Number of items in each domain	Minimum number of tests to fail to be considered impaired on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

0 Domain unaffected. Represents performance below cut-off on one or less tests in that domain.

1 Domain impaired. Represents performance below cut-off on two or more tests in the domain.

Total domain impairments	Verbal memory	Visuospatial memory	Processing speed	Visuo-perceptual	Language	Number of pwMS
No impairments	0	0	0	0	0	27
Impairment on only one domain	1	0	0	0	0	3
	0	1	0	0	0	3
	0	0	1	0	0	7
	0	0	0	1	0	2
	0	0	0	0	1	
Impairment on two domains	1	1	0	0	0	
	1	0	1	0	0	3
	1	0	0	1	0	2
	1	0	0	0	1	
	0	1	1	0	0	
	0	1	0	1	0	1
	0	1	0	0	1	
	0	0	1	1	0	3
	0	0	1	0	1	
	0	0	0	1	1	
Impairment on three domains	1	1	1	0	0	1
	1	1	0	1	0	
	1	1	0	0	1	
	1	0	1	1	0	8
	1	0	1	0	1	
	1	0	0	1	1	
	0	1	1	1	0	
	0	1	1	0	1	
	0	1	0	1	1	
	0	0	1	1	1	
Impairment of four domains	1	1	1	1	0	1
	1	1	1	0	1	
	1	1	0	1	1	
	1	0	1	1	1	
	0	1	1	1	1	
Global impairment	1	1	1	1	1	

This finding has raised a question of whether these individuals could be considered to represent a pattern of MS progression – it could be hypothesized that globalization of MS cognitive deficits starts with problems in either verbal memory, processing speed or visuoperceptual cognition; then with longer disease duration spreads into all three domains; and with even more progression and longer MS duration impairments spread into visuospatial memory.

In order to address this question I looked into the MS types, disease duration and neurological disability scores of these nine patients. The patient with 4-domain impairments was a 52-year-old female with SPMS and 9.4 years MS duration. Of the eight patients who had failed three cognitive domains the majority (five) had SPMS, one had PPMS, and two had RRMS. In addition, all of these patients have had long MS duration (18.63 (± 11.25) years), and most of them were older (54.66 (± 10.39) years) and had more neurological disability (5.72 (± 1.46) EDSS score).

One particular finding that can be noted was that at phase II none of the patients had failed the language domain, even though eight patients had failed the TROG and two patients failed the BC.Index.

In addition, at phase II the impairments on the processing speed domain became more evident, and among the single-domain cognitive deficits, processing speed impairments became around twice more frequent than any other cognitive domain impairments.

5.3.3. Dimensionality of temporal change in cognitive impairments

In the previous analyses the patterns were shown of how cognitive impairments group in pwMS at two separate phases of the assessment. However, what was left unexplained was how stable those impairments were, and what was the pattern of their longitudinal development.

The dimensionality of change was analysed in two ways and separate steps were taken to investigate the impairments that emerge, and the impairments that resolve. This was done in order to explore whether any patterns of domain-specificity of evolving or resolving impairments could be observed.

Several measures were undertaken in order to avoid the effect of potential artefacts in the dimensionality analysis that could potentially limit the generalizability of the findings. I identified a possibility that there could also be individuals that show patterns of both emerging and resolving

impairments in the same domain, but on different tests. Because of this reason it was chosen to define longitudinal change on a cognitive domain when there were at least two observations on two separate tests in the same direction, following the same reasoning as per definition of impairment in Table 5.3 and Table 5.4. This adjustment made the analyses of longitudinal change more conservative, especially on the domains that have fewer tests in them. This method has made the chances of making Type I error smaller, but at the same time inflated the chances of the Type II error.

I. Dimensionality of emerging impairments

The first part of longitudinal analyses of changes in dimensionality of cognitive impairment revolved around identification of evolving impairments. These were the instances where a patient has performed at norm at phase I, but below the cut-off at phase II (Table 5.5). The majority of patients did not develop any new domain impairments, at least none that could get picked up by this conservative definition. Almost all of the newly developed domain impairments were uni-dimensional, with emerging impairments in processing speed and visuoperception most prominent. The high frequency of new impairments in the visual cognition domain could be potentially an artefact arising from the fact that this domain contained the highest number of cognitive tests, therefore the probability of acquiring new deficits in two of them was the highest. In addition, none of the patients had developed new deficits in the language domain.

Table 5.5. Dimensionality of developing impairments. The pattern of acquired deficits for pwMS (n=61) in each domain between phase I and phase II is presented.

Domain	Number of tests in each domain	Minimum number of items to develop new impairments on to be considered having emerging impairments on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

0 Impairments resolved or remained stable.

1 New impairments that emerged.

Total new domain impairments	Verbal memory	Visuospatial memory	Processing speed	Visuo-perceptual	Language	Number of pwMS
No new impairments	0	0	0	0	0	39
New impairment on only one domain	1	0	0	0	0	2
	0	1	0	0	0	2
	0	0	1	0	0	6
	0	0	0	1	0	7
	0	0	0	0	1	
New impairment on two domains	1	1	0	0	0	
	1	0	1	0	0	1
	1	0	0	1	0	3
	1	0	0	0	1	
	0	1	1	0	0	
	0	1	0	1	0	
	0	1	0	0	1	
	0	0	1	1	0	1
	0	0	1	0	1	
	0	0	0	1	1	
New impairment on three domains	1	1	1	0	0	
	1	1	0	1	0	
	1	1	0	0	1	
	1	0	1	1	0	
	1	0	1	0	1	
	1	0	0	1	1	
	0	1	1	1	0	
	0	1	1	0	1	
	0	1	0	1	1	
	0	0	1	1	1	
New impairment on four domains	1	1	1	1	0	
	1	1	1	0	1	
	1	1	0	1	1	
	1	0	1	1	1	
	0	1	1	1	1	
Sudden global impairment	1	1	1	1	1	

Table 5.6. Dimensionality of resolving impairments. The pattern of resolving deficits for pwMS (n=61) in each domain between phase I and phase II is presented.

Domain	Number of tests in each domain	Minimum number of tests to improve on to be considered having resolving impairments on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

1 Impairments resolve. I.e. impaired at phase I and non-impaired at phase II

0 New impairments that emerge or remain stable.

Total domain recovery	Verbal memory	Visuospatial memory	Processing speed	Visuo-perceptual	Language	Number of pwMS
No recovery	0	0	0	0	0	51
Recovery on only one domain	1	0	0	0	0	4
	0	1	0	0	0	1
	0	0	1	0	0	
	0	0	0	1	0	2
	0	0	0	0	1	1
Recovery on two domains	1	1	0	0	0	
	1	0	1	0	0	
	1	0	0	1	0	2
	1	0	0	0	1	
	0	1	1	0	0	
	0	1	0	1	0	
	0	1	0	0	1	
	0	0	1	1	0	
	0	0	1	0	1	
	0	0	0	1	1	
Recovery on three domains	1	1	1	0	0	
	1	1	0	1	0	
	1	1	0	0	1	
	1	0	1	1	0	
	1	0	1	0	1	
	1	0	0	1	1	
	0	1	1	1	0	
	0	1	1	0	1	
	0	1	0	1	1	
	0	0	1	1	1	
Recovery on four domains	1	1	1	1	0	
	1	1	1	0	1	
	1	1	0	1	1	
	1	0	1	1	1	
	0	1	1	1	1	
Global recovery	1	1	1	1	1	

5.3.4. Dimensionality of resolving impairments

The vast majority of patients did not improve on any of the cognitive domains, and if they did, they mostly improved on only one domain (Table 5.6.). Therefore it could be considered, that in general, the changes between phases I and II were not random, instead, a trend for acquiring new impairments was observed, with most of new problems emerging on visuoperceptual and information processing speed tests.

5.3.5. Dimensionality of cognitive impairments. Summary

From analysing the domain-specificity of cognitive impairments it was found that all cognitive domains can be affected in MS, although not at the same frequencies. The majority of individuals from the observed MS sample had deficits only on one cognitive domain, however, with when re-assessed three years later they became impaired on more cognitive domains, with emerging deficits on visuoperceptual and information processing speed tests being most prominent.

5.4. The dimensionality of the study cohort

After having analysed how the impairments of different cognitive domains group together, the next step was to investigate the pattern of *how the pwMS from our sample could be grouped* according to their cognitive performance. In this section it was aimed to examine, without making any predictions, whether any patterns could be observed between the emerging groups of patients in respect to their cognitive difficulties.

5.4.1. The dimensionality of the phase I study population

It was chosen to run a cluster analysis in order to address the emerging patterns of grouping the pwMS based on their performance. The investigation of dimensionality of the phase I cohort was started by determining the most appropriate number of clusters for the following classifications.

I. Number of subject-clusters in the phase I dataset

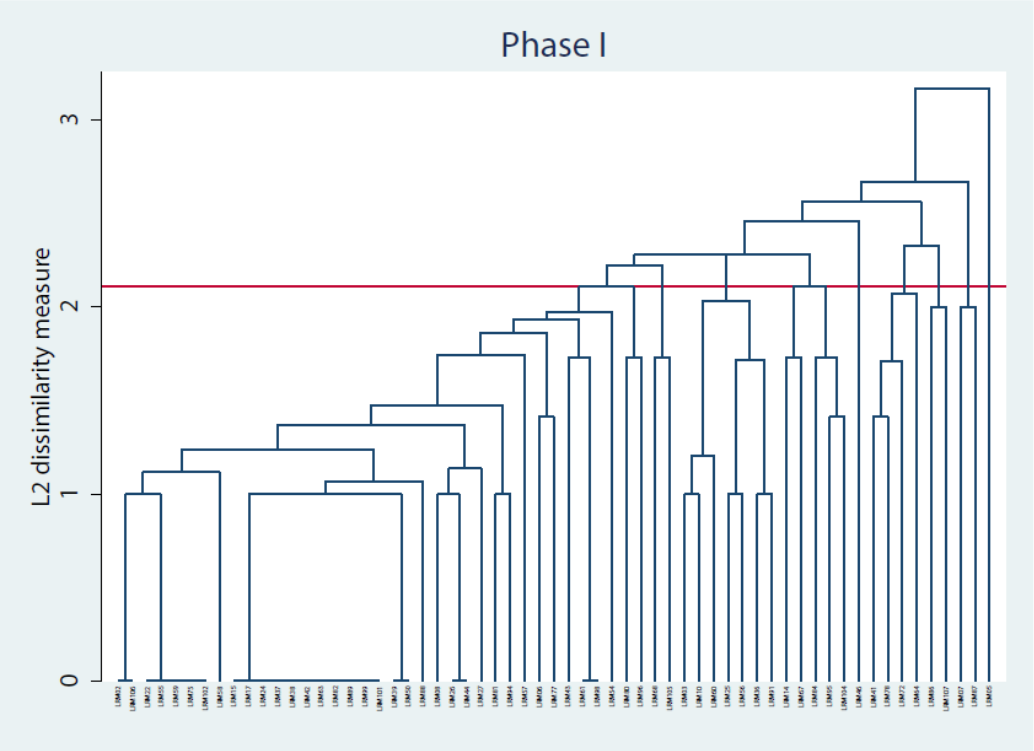
Based on the Caliński & Harabasz F-values (Table 5.7) a ten-cluster solution provided the best summary of the dataset. As it can be seen from the dendrogram (Figure 5.1), the ten clusters vary greatly in their sizes. The raw data of performance of individuals in each cluster can be seen in Appendix F.

Table 5.7. Determination of the optimal number of clusters in the hierarchical cluster analysis of patient data at phase I

Number of Clusters	Caliński-Harabasz pseudo-F
2	3.39
3	3.59
4	5.28
5	4.6
6	4.22
7	4.72
8	6.22
9	6.29
10	**6.33**
11	6.15
12	5.9
13	6.1
14	5.65

Table 5.7. indicates that the Calinski-Harabasz pseudo-F supports a ten cluster solution.

Figure 5.1. Dendrogram of hierarchical cluster analysis using average linkage model of patient data at phase I



Note. The horizontal line is shown to represent where the 10-cluster cut falls in this analysis

II. Neuropsychological characteristics of pwMS in each subject-cluster in phase I

The frequency of cognitive impairments for each test by subject-cluster is shown in Table 5.8. An interpretative view of the cluster solution follows:

The largest cluster (C1: 35/61 = 57.4%) showed impairments that appear to occur scattered across cognitive abilities, consistent with the classic pathological view of a widely distributed multifocal inflammatory disorder of the CNS.

Table 5.8. Frequencies (%) of patients failing each test in each cluster at phase I

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
10-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	35	0	6	14	6	9	6	6	3	0	3	29	9	3	3	6	31
C2	2	50	0	0	0	0	**100**	0	50	0	50	0	0	0	0	**100**	0
C3	2	0	0	**100**	50	0	**100**	0	0	**100**	0	0	50	0	0	0	50
C4	7	**100**	**100**	**71**	43	0	43	0	0	0	14	57	0	29	0	0	14
C5	5	**80**	40	**100**	0	40	20	0	20	0	20	0	**100**	0	0	20	40
C6	1	0	0	0	0	0	0	0	**100**	**100**	**100**	0	0	0	0	**100**	**100**
C7	4	**75**	**100**	50	0	0	**100**	**75**	**100**	0	50	25	0	0	0	0	50
C8	2	50	0	50	0	0	**100**	**100**	**100**	0	0	**100**	0	0	50	**100**	50
C9	2	50	0	50	0	**100**	0	0	50	0	50	0	0	**100**	**100**	**100**	0
C10	1	**100**	**100**	**100**	**100**	**100**	**100**	0	**100**	**100**	0	0	**100**	**100**	**100**	0	0
Total	**61**																
Average		**51**	**35**	**54**	**20**	**24**	**57**	**18**	**52**	**30**	**29**	**21**	**26**	**23**	**25**	**43**	**34**

C1: Sporadic impairments with possible slight excess of visuoperceptual & language impairments (VOSP6 & BC.Index)
C2: Processing speed and grammar reception impairment (SDMT & TROG)
C3: Verbal recall, processing speed and naming impairments (SRTD, SDMT & VOSP 2)
C4: Verbal learning and memory (with or without visual cognition impairments) (SRT)
C5: Verbal memory and counting (SRT & VOSP 8)
C6: Language and naming impairments (WLGT, VOSP 2 & 3, TROG & BC.Index)
C7: Verbal memory and processing speed (SRT, SDMT, PASAT, WLGT)
C8: Processing speed, position discrimination and grammar reception (SDMT, PASAT, WLGT, VOSP6 & TROG)
C9: Space perception (SPARTD, BORB 2 & 5) and grammar reception (TROG)
C10: Global impairment

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC– Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

Cluster two (C2: 2/61 = 3.3%) indicated impairments on one of the processing speed tests and one language test, and cluster three (2/61 = 3.3%) patients had impairments on one of processing speed tests, verbal recall and naming. Since in these clusters the patients failed only one test per domain but not the others, the clinical validity of such classification was unclear and these clusters could be considered either artefacts, or parts of cluster one.

Cluster four (7/61 = 11.5%) was one of the clearest clusters and the patients in it had a core impairment of verbal memory.

Cluster five (5/61 = 8.2%) grouped patients with difficulties counting cubes and remembering words, with or without additional impairments. The value of this distinction is also unclear.

The individual in cluster six (1/61 = 1.6%) had a clear impairment of language and failed only the tasks of word and sentence generation, grammar comprehension and picture naming.

Cluster seven (4/61 = 6.6%) individuals had a core impairment of verbal memory and processing speed. The four individuals in this cluster fit the cognitive profile of impairment measured by the BRBN battery.

Cluster eight (2/61 = 3.3%) represented individuals who had impairments in processing speed, position discrimination, and reception of grammar. It may be that clusters eight and seven were parts of the same entity that represents individuals with processing speed impairments with or without additional problems (possibly the additional problems that are caused or mediated by impairments of processing speed).

Cluster nine (2/61 = 3.3%) represented individuals with difficulties in space perception with an additional impairment of grammar reception.

Cluster ten (1/61 = 1.6%) contained an outlier case with impairments on multiple tasks across all cognitive functions.

III. Demographic and clinical characteristics of pwMS in each subject-cluster in phase I

The demographic and clinical data of patients in each cluster is presented in Table 5.9.

From inspecting the characteristics of individuals in cluster one, no clear excess of specific clinical or demographic variables emerged. Patients in this cluster therefore appeared to represent the typical MS cohort with the classic pathological view of a widely distributed multifocal inflammatory disorder of the CNS.

Table 5.9. Demographic and clinical characteristics of pwMS in each subject-cluster in phase I

Test	C1 (n=35)	C2 (n=2)	C3 (n=2)	C4 (n=7)	C5 (n=5)	C6 (n=1)	C7 (n=4)	C8 (n=2)	C9 (n=2)	C10 (n=1)
Age	46.66 (9.25)	**38.5 (9.19)**	**51.5 (13.44)**	46.71 (11.8)	48.4 (8.02)	**36**	45.75 (11.67)	52 (11.31)	52.5 (3.54)	49
Gender (% female)	49%	50%	50%	57.1%	40%	0%	75%	50%	50%	100%
Years of education	14.32 (2.98)	**10 (0)**	13 (5.65)	10.86 (1.57)	12.2 (3.03)	**10**	11.5 (3.42)	10.5 (0.71)	11 (0)	10
Premorbid ability (NART score)	117.23 (7.9)	**92.5 (17.68)**	113 (0)	114.57 (4.58)	113.8 (9.01)	**86**	107.75 (8.46)	100 (7.07)	117 (5.66)	100
Dementia (ACE-R <82)	0%	0%	0%	0%	0%	**100%**	25%	0%	0%	100%
BDI-II depression score	15.65 (9.42)	19.5 (16.26)	11 (4.24)	15.14 (8.01)	15.4 (10.26)	**29**	14 (16.6)	9 (8.48)	9 (5.66)	22
MS disease course	26% PPMS, 23% SPMS, 51% RRMS	**100% SPMS**	**100% RRMS**	29% PPMS, 29% SPMS, 42% RRMS	**40% PPMS, 60% RRMS**	100% RRMS	25% PPMS, 25% SPMS, 50% RRMS	50% PPMS, 50% SPMS	50% PPMS, 50% SPMS	100% SPMS
Disease duration (years)	8.57 (5.18)	9.5 (12.02)	**5 (2.83)**	12.71 (10.13)	10.8 (4.15)	7	12.25 (15.48)	9 (9.49)	10.5 (7.78)	6
EDSS score	4.06 (2.3)	6 (0)	**2 (0)**	4.43 (1.62)	4.4 (1.67)	2	5 (2)	6 (0)	6 (0)	6
MS modifying treatment	23%	0%	50%	28.6%	20%	100%	25%	0%	0%	100%
Antidepressant use	31%	0%	50%	14.3%	20%	0%	**100%**	50%	50%	100%

Note. The descriptive data presented in this table was collected at phase I
Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, NART – National Adult Reading Test, BDI-II – Beck's Depression Inventory 2nd Ed., EDSS – Expanded Disability Status Scale

Although clusters two and three did not seem to be much different in terms of cognition, and in the previous section I speculated that they might even be from the same entity, differences were seen in their clinical characteristics. Cluster two grouped SPMS patients who were younger, less educated and more depressed, and cluster three grouped RRMS patients that were older with less neurological disability and shorter MS duration. This could be used to argue that if clusters one, two and three represented the same group in terms of cognition, then the patients from clusters two and three could be considered being at the opposite ends of the spectrum in terms of demographic and clinical characteristics, with cluster one in-between.

Cluster four appears to include all types of patients and there seemed to be no specific excess of clinical or demographic variables. Similarly, cluster five did not show clear excess for any clinical or demographic characteristics. These individuals may belong to cluster four and thus could be grouped as patients with core verbal memory difficulties with additional (potentially unrelated) impairments in spatial awareness or other domains.

The language impairments and low performance on the ACE-R test of the individual in cluster six could potentially be caused by his low education, and not factors related to MS.

The cognitive impairments (verbal memory and processing speed) of patients in cluster seven could potentially be side-effects of their antidepressant medication. By not finding any major differences in terms of clinical or demographic variables between the individuals in clusters seven and eight, the possibility remains that that these two clusters form parts of the same entity and represent core impairments of processing speed potentially caused by anti-depressant uptake.

The two patients in cluster nine seem to exhibit cognitive impairments that could not be explained by clinical or demographic factors, and thus potentially could be indicative of posterior lesions or posterior atrophy.

Cluster ten represents an outlier case with aggressive MS. At phase I this woman with SPMS has had MS for only six years, but her neurological disability and cognitive dysfunction had progressed profoundly and she seemed to be doing much worse than other patients.

IV. The dimensionality of Phase I study population. Summary

From analysing the dimensionality of the phase I cohort it was found that the MS population is highly heterogeneous in respect to the dimensionality of their cognitive impairments. Speculations could be made about the best ways to group the patients, however, they do not yield a clear answer.

5.4.2. The dimensionality of the phase II study population

After exploring the dimensionality of the phase I cohort the next step was to investigate whether the cluster solution for the population dataset of impairments simplified three years later at phase II. This in turn would answer the question whether our MS population starts as highly heterogeneous and then gradually becomes more homogeneous over time. Therefore the same cluster analysis methodology was therefore repeated on the same 61 pwMS at phase II.

I. Number of subject-clusters in the phase II dataset

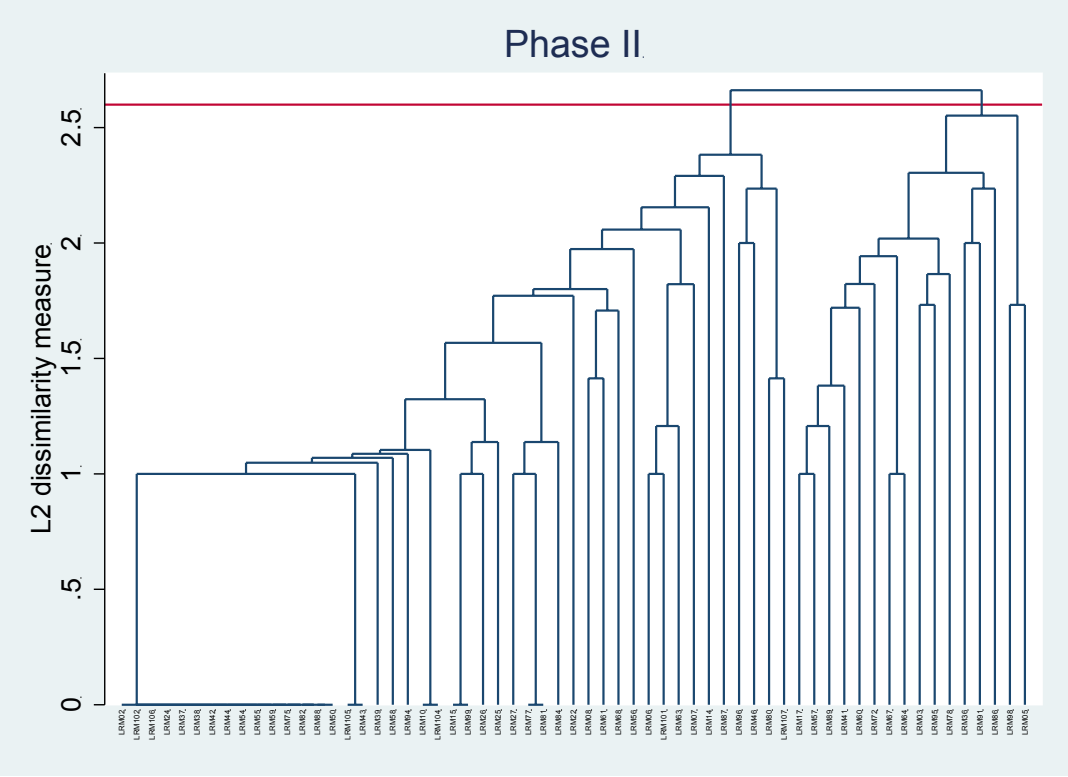
The number of clusters for grouping the phase II was selected by examining the values of the Calinski-Harabasz pseudo-F statistic. As it can be seen from Table 5.10, the F-statistic was the highest when the sample was broken down into two clusters. Therefore the two cluster model was chosen to employ for grouping the phase II data (Figure 5.2).

Table 5.10. Determination of the optimal number of clusters in the hierarchical cluster analysis of patient data at phase II

Number of Clusters	Caliński-Harabasz pseudo-F
2	**23.01**
3	14.06
4	12.54
5	10.7
6	9.39
7	-
8	7.65
9	7.22
10	8.19
11	8.04
12	-
13	7.23
14	7.16

Table 5.10. indicates that the Calinski-Harabasz pseudo-F supports a two cluster solution.

Figure 5.2. Dendrogram of hierarchical cluster analysis using average linkage model of patient data at phase II.



Note. The horizontal line is shown to represent where the 2-cluster cut falls in this analysis

II. Neuropsychological characteristics of pwMS in subject-clusters at phase II

The frequencies of cognitive impairments for each test by subject-cluster are shown in Table 5.11 (individual performance of each patient can be seen in Appendix F). An interpretative view of the cluster solution follows:

The largest cluster (C1: 45/61 = 73.8%) showed impairments that appeared to occur scattered across cognitive abilities, consistent with the classic pathological view of a widely distributed multifocal inflammatory disorder of the CNS.

Cluster two (C2: 16/61 = 26.2%) showed core impairments of processing speed and verbal memory with or without visuoperceptual impairment.

Table 5.11. Frequencies (%) of patients failing each test in each subject-cluster at phase II

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Langu-age	
10-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	45	11	4	13	9	16	27	31	7	9	9	4	11	4	16	9	4
2	16	100	88	88	31	13	100	88	6	31	44	50	38	25	31	25	0
Total 61																	
Average		**56**	**46**	**51**	**20**	**15**	**64**	**60**	**7**	**20**	**27**	**27**	**25**	**15**	**24**	**17**	**2**

C1: Sporadic impairments (with possible slight excess of processing speed impairments)
C2: Verbal memory and processing speed impairments with or without impairments in visuoperceptual functions

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC– Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

III. Demographic and clinical characteristics of pwMS in each subject-cluster at phase II

The analysis of demographic and clinical characteristics of the patients from the two clusters is presented in Table 5.12. The main difference between the two clusters was that cluster one had more RRMS, and cluster two had more SPMS patients in them. Naturally, consistent with the differences between these two MS types, the patients in cluster two had longer MS duration, although this difference did not reach statistical significance. The patients in cluster two also tend to have higher frequencies of being on antidepressant medication.

Table 5.12. Demographic and clinical characteristics of pwMS in each subject-cluster at phase II.

Test	C1 (n=45)	C2 (n=16)	Difference
Age	49.39 (9.32)	52.27 (9.06)	U=267, p=0.272
Gender (% female)	47%	63%	p = 0.243
Years of education	13.52 (3.2)	11.93 (2.76)	U=225.5, p=0.102
Premorbid ability (NART score)	114.2 (10.44)	112.31 (8.86)	U=304.5, p=0.362
Dementia (ACE-R <82)	7%	25%	p = 0.062
BDI-II depression score	14.57 (9.58)	17.73 (8.42)	U=235.5, p=0.099
MS disease course	**29% PPMS, 16% SPMS, 55% RRMS**	**19% PPMS, 56% SPMS, 25% RRMS**	**p=0.006**
Disease duration (years)	11.6 (4.72)	16.13 (10.32)	U=254.4, p=0.226
EDSS score	4.98 (1.91)	5.47 (1.51)	U=293, p=0.494
MS modifying treatment	53%	56%	p = 0.771
Antidepressant use	**29%**	**56%**	**p = 0.031**

Note. The descriptives presented in this table were collected at phase II, except for the NART, which was collected at baseline. The between-group comparisons were performed using the Mann-Whitney U-test for interval data and Fisher's exact test for frequency data

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, NART – National Adult Reading Test, BDI-II – Beck's Depression Inventory 2nd Ed., EDSS – Expanded Disability Status Scale

IV. The dimensionality of phase II study population. Summary

From analysing the dimensionality of the phase II cohort it appeared that the heterogeneity of the cohort in respect to their cognitive impairment diminishes with time, and the profiles of cognitive performance tend to become more homogeneous at phase II.

5.4.3. The dimensionality of the study population with respect to longitudinal change in cognitive impairments

As it can be seen from the analyses in section 5.3.1. and section 5.4.2. earlier in this chapter, the cluster analyses have yielded different results when employed at phase I and at phase II. From this it can be assumed that there have been time-related changes in the dimensionality of the study population. In the following section the temporal changes in MS-related cognitive deficits were examined by using the cluster analysis to group the patients based on the trajectory of their longitudinal changes. This was carried out to identify individuals whose cognitive deficits have progressed, remained the same, or improved over the period of three years.

I. Number of clusters

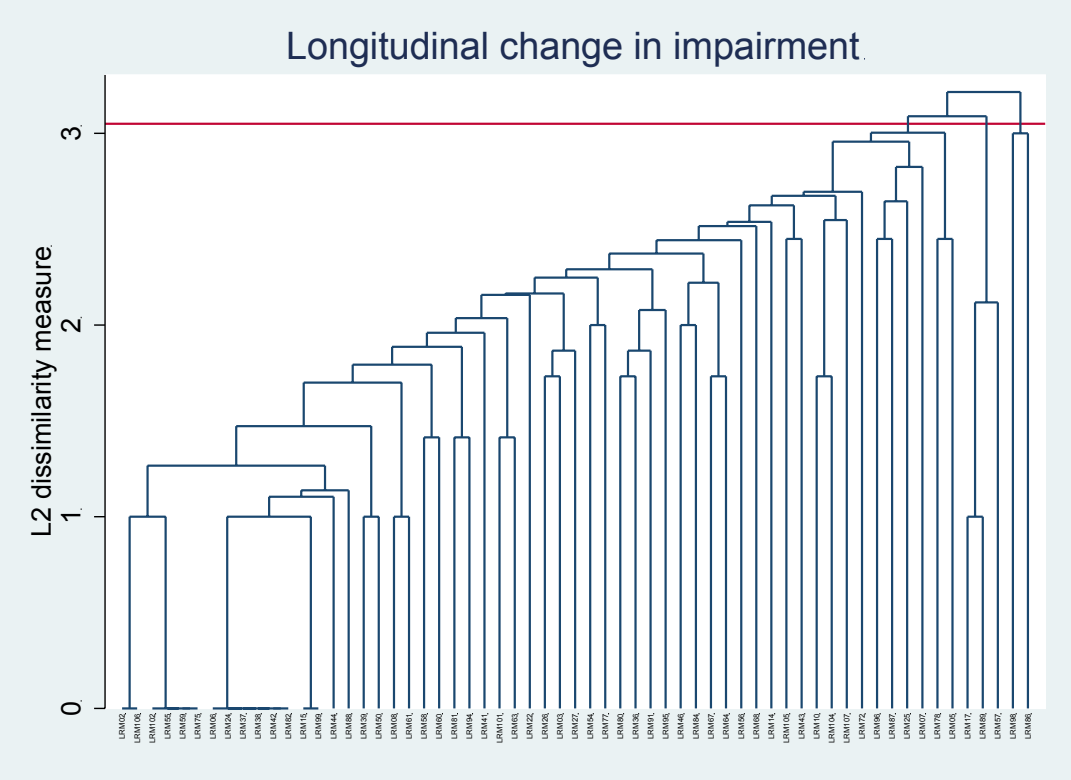
The cluster analysis was employed in order to group the pwMS based on their longitudinal changes in cognitive test performance. Based on the F-statistic (Table 5.13) we found that four trajectories of longitudinal change in cognitive performance could be identified (Figure 5.3 & Table 5.14).

Table 5.13. Determination of the optimal number of clusters in the hierarchical cluster analysis of patient longitudinal change in their cognitive performance between phases I and II

Number of Clusters	Calinski-Harabasz pseudo-F
2	2.4
3	4.06
4	**4.23**
5	3.66
6	4.22
7	3.84
8	3.63
9	3.83
10	3.61
11	3.65
12	3.52
13	3.46
14	3.41

Table 5.13. indicates that the Calinski-Harabasz pseudo-F supports a four cluster solution.

Figure 5.3. Dendrogram of hierarchical cluster analysis model of longitudinal trajectories of change in patient performance on cognitive tests between phases I and II



Note. The horizontal line is shown to represent where the cluster cut falls in this analysis

Table 5.14. Neuropsychological characteristics for the four subject-cluster solution for longitudinal change. Frequencies (%) of three different trajectories of change on the cognitive tests are presented

		Resolved impairment	Unchanged	New impairment
CLUSTER ONE (n=54)	SRTL	13	78	9
	SRTC	11	87	2
	SRTD	17	79	4
	SPART	6	88	6
	SPARTD	7	84	9
	SDMT	2	81	17
	PASAT	0	69	31
	WLGT	15	79	6
	VOSP2	4	87	9
	VOSP3	7	84	9
	VOSP6	24	70	6
	VOSP8	13	74	13
	BORB2	6	90	4
	BORB5	0	87	13
	TROG	9	91	0
	BC.Index	33	63	4

		Resolved impairment	Unchanged	New impairment
CLUSTER TWO (n=3)	SRTL	0	67	33
	SRTC	0	67	33
	SRTD	0	67	33
	SPART	0	33	67
	SPARTD	0	100	0
	SDMT	0	100	0
	PASAT	0	67	33
	WLGT	67	33	0
	VOSP2	0	67	33
	VOSP3	0	33	67
	VOSP6	0	33	67
	VOSP8	0	67	33
	BORB2	33	67	0
	BORB5	0	67	33
	TROG	0	0	100
	BC.Index	0	100	0

Table 5.14. (continued)

		Resolved impairment	Unchanged	New impairment
CLUSTER THREE (n=3)	SRTL	0	0	100
	SRTC	0	0	100
	SRTD	0	0	100
	SPART	0	100	0
	SPARTD	0	100	0
	SDMT	0	0	100
	PASAT	0	0	100
	WLGT	0	100	0
	VOSP2	0	100	0
	VOSP3	0	100	0
	VOSP6	0	33	67
	VOSP8	33	33	33
	BORB2	0	67	33
	BORB5	33	67	0
	TROG	0	100	0
	BC.Index	33	67	0

		Resolved impairment	Unchanged	New impairment
CLUSTER FOUR (n=1)	SRTL	0	0	100
	SRTC	0	0	100
	SRTD	0	0	100
	SPART	0	100	0
	SPARTD	0	100	0
	SDMT	0	100	0
	PASAT	0	100	0
	WLGT	100	0	0
	VOSP2	0	0	100
	VOSP3	0	100	0
	VOSP6	100	0	0
	VOSP8	0	0	100
	BORB2	0	0	100
	BORB5	0	100	0
	TROG	0	100	0
	BC.Index	0	100	0

C1: New impairments develop on tasks of processing speed (SDMT and PASAT) and one of the visuoperceptual tests (BORB5). Practice effect on many tasks (especially language) is observed.
C2: Progression of deficits in grammar reception (TROG) with or without additional deficits
C3: Progression of deficits in verbal memory and processing speed with or without progression of deficits in spatial awareness
C4: Progression of deficits in verbal memory and visuoperceptual tests

I. Neuropsychological characteristics longitudinal change in subject-clusters of pwMS

Individual changes in subject-clusters for each patient can be seen in the Appendix F. An interpretative view of the cluster solutions follows:

The largest cluster (C1: 54/61 = 88.5%) shows a pattern of impairments emerging or resolving across cognitive abilities, although a tendency for emerging impairments in information processing speed is observed. A practice effect is also observed, but mainly on tasks involving language (sentence and word generation, reception of grammar, and verbal memory).

Cluster two (C2: 3/61 = 4.9%) shows emergence of impairments in grammar reception (TROG) with or without additional deficits.

Cluster three (C3: 3/61 = 3.3%) shows emergence of impairments in verbal memory and processing speed with or without progression of deficits in spatial awareness.

Cluster four (C4: 1/61 = 1.6%) shows an individual whose deficits progressed in verbal memory and visuoperceptual cognition.

II. Demographic and clinical characteristics of pwMS in each subject-cluster

The demographical and clinical descriptives of the individuals from each cluster representing the four trajectories of longitudinal change can be seen in Table 5.15. From inspecting the characteristics of individuals in cluster one, no clear excess of specific clinical or demographic variables emerged. This group therefore appeared to represent a typical cohort of pwMS.

Clusters two, three and four could be considered to form one cluster of changes relevant only to patients with SPMS. It could be considered, that the cognitive changes experienced by patients with SPMS could potentially be different from those experienced by all other pwMS.

Table 5.15. Demographic and clinical characteristics of pwMS from each of the four subject-clusters representing the four trajectories of longitudinal change between performance in phase I and phase II

Test	C1 (n=54)	C2 (n=3)	C3 (n=3)	C4 (n=1)
Age	49.8 (9.38)	54.33 (6.8)	52.33 (12.42)	47
Gender (% female)	50%	67%	67%	100%
Years of education	13.26 (3.17)	11 (3.6)	13.67 (2.52)	10
Premorbid ability (NART score)	114.26 (9.77)	105 (11.36)	118.67 (3.79)	95
Dementia (ACE-R <82)	7%	67%	0%	100%
BDI-II depression score	14.54 (9.32)	22.67 (5.51)	17.33 (7.64)	31
MS disease course	29% PPMS, 17% SPMS, 54% RRMS	100% SPMS	100% SPMS	100% SPMS
Disease duration (years)	12.22 (5.96)	23.67 (14.5)	9.67 (3.79)	18
EDSS score	4.94 (1.85)	5.67 (0.57)	6.67 (1.16)	7
MS modifying treatment	54%	67%	67%	0%
Antidepressant use	32%	67%	67%	0%

Note. The descriptives presented in this table were collected at phase II, except for the NART, which was collected at phase I

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, NART – National Adult Reading Test, BDI-II – Beck's Depression Inventory 2nd Ed., EDSS – Expanded Disability Status Scale

III. Longitudinal changes in the dimensionality of the study population. Summary

Four clusters of longitudinal change in the dimensionality of cognitive impairments were observed, and the patients developed new deficits or showed improvement in all cognitive domains. Patients with SPMS appeared to have their cognitive deficits progress in a different manner than the rest of the patients.

5.5. Chapter summary

In Chapter Five from analysing the dimensionality of cognitive impairments in MS, several tendencies have emerged.

First, the cognitive deficits were seen to be prevalent in all cognitive domains and tests (none were spared), and initially they appeared heterogeneous on individual basis, although with MS progression, and development of new impairments in information processing speed and visuoperceptual functions, had the tendency to become more homogenous.

Second, there was also a reduction in the dimensionality of the MS population with respect to their pattern of cognitive impairments over time. At phase I the patients classified into multiple groups based on their performance, indicative of the many ways in which cognitive deficits start. However, three years later (at phase II), only two clusters were present; this showed that with acquirement of new deficits the patients became more homogeneous in the expression of their cognitive symptoms.

The analyses performed in this chapter would imply that MS affects individuals by initially causing a small set of domain-specific impairments (that differ between patients), and later evolves into (a larger) multi-domain set of impairments (that has more commonalities between patients). Most commonly the new deficits were in the information processing speed domain, however, besides that there seemed to be no clear pattern for new deficit acquirement, as it appeared that the deficits that the patients picked up have been sporadic. Our analyses suggest that in MS the cognitive deficits slowly accumulate, leading to development of major problems with longer disease duration.

Through these analyses we were unable to identify a group of individuals who suddenly flipped into progression and whose cognitive functioning had changed globally. Instead the analyses indicated that the cognitive deficits have been acquired (and lost) mostly at random (with only very slight predisposition towards new impairments on information processing speed). Moreover, it could be speculated that some individuals with SPMS may have a slightly different course of progression of cognitive deficits, but the results do not support any definite conclusions regarding differences in the dimensionality of cognitive impairments among different subtypes of MS.

Chapter Six. The trajectory, extent, and predictors of cognitive change in pwMS

6.1. Chapter overview

From analysing the dimensionality of cognitive impairment in chapter five I found that during the span of 3 years each patient has improved on some tests, and deteriorated on the others. However, the previous chapter investigated only dimensionality without addressing the magnitude of cognitive change in each domain. This chapter was dedicated to the investigation of the longitudinal change in cognition associated with MS, and identification of factors that are related to better or poorer cognitive outcome.

The chapter begins with an analysis of distribution of controls' scores at phases I and II of the study in search for systematic differences in changes of performance that could be attributable to variables not related to MS. Then the main analysis revolved around the variation of patients' change in each cognitive domain, and the extent to which the change could be predicted by clinical and demographic variables, and cognitive reserve.

6.2. Methods

6.2.1. Participants

For the analyses of the trajectory and extent of longitudinal change the data from the participants who underwent both phase I and phase II assessments was used (23 controls and 82 pwMS). However, for further analyses that required averaging the raw scores on the tests in order to establish standardized scores of domain-specific change, only the performance of pwMS with full data on all tests (n=61) was used.

6.2.2. Procedure

I. The trajectory and extent of cognitive change

Prior to running any analyses on the change in patients' performance, I needed to have an estimation of how much variation in longitudinal change is normal. This approach takes into account that not only patients change in their performance, but controls do as well, and that age-related deterioration on certain tasks, and practice effect on others were expected. Moreover, it was acknowledged that because the phase I and phase II assessments were performed by separate researchers, there was a chance for systematic differences in administration. Therefore in order to control for all these sources of normal variation I started this chapter by analysing the change in performance exhibited by healthy controls.

a) Control change

The comparisons of control means between the performances at phase I and at phase II were presented in the previous chapter, Table 5.1. From that table we could see how much the controls have changed on each individual test, but the amount of change in raw numbers didn't allow true appreciation of the change in one test in relation to another. For example, a change in 1 point on VOSP 8 (where the maximum score is 10 points) represented a much larger change than a change in 1 point on BORB 5 (where the maximum score is 40 points). Therefore to estimate the true magnitude of how much the control performance varied from phase I to phase II on each test, the raw scores on all of the test items were converted into percentages. This conversion also allowed a better appreciation of the relative difficulty of each cognitive test, as the tests that were difficult for all participants would have lower values when converted into percentages, and easy tests with ceiling effect where the values approached 100% would also be easier to identify.

First the difference between the performance on the phase I and phase II tests was calculated for the controls by subtracting phase I performance percentage from the phase II performance percentage. By calculating the performance differences this way, it was made sure that a positive difference

between the baseline and follow-up represented improvement, while a negative difference represented deterioration in performance.

b) Patient change

After analysing the trajectory and extent of the control change, the next step was to investigate how much the patients have changed in the three years between the phase I and phase II assessments. Again, this procedure was performed after converting the raw data into percentages to allow for comparison of the extent of change among the different tests. In this chapter I aimed to analyse not only the extent of how much the patients have changed, but also the trajectory of their change, since I expected both decline (due to MS progression) and improvement (due to practice effect) to be observed. The patient change was calculated in the same way as the control change, i.e. by subtracting phase I performance percentage from the phase II performance percentage.

c) Comparison of the trajectory and extent between control and patient change

The change in performance between control and patient participants was compared using the Mann-Whitney U-test. A non-parametric test was chosen to employ as ceiling effect was expected on some of the cognitive tests, resulting in little variation in performance.

II. *Standardization of change*

The natural tendency to experience changes in performance due to factors not related to MS needed to be taken into account later on when analysing the patient data in order to avoid overestimations or underestimations of deterioration. For example, on the tests where the controls performed worse at follow-up extra caution was needed for not overestimating the decline, as part of the decline observed in patients may be caused by other than MS-related factors. In addition, on tests where the controls showed practice effect, extra caution was needed for not underestimating the impairments. For example, if a patient appeared not to have changed, this could potentially indicate that the MS-related decline may be masked by practice effect, and the real change might be underestimated. These considerations have led to justification of a novel approach for further data analysis in order to adjust for factors related to healthy cognitive ageing, practice effect, and systematic differences in

administration. By running the following procedures it was attempted to account for those systematic effects.

The z-score approach was chosen to employ since standardizing the variables allows comparing each patient outcome with that found in the reference control population. This enabled to account for normal variation in performance, expressed by normal deterioration related to cognitive ageing, and to improvement, related to practice effect. In addition, by choosing to convert the raw performance scores to z-scores I attempted to adjust for systematic bias in administration, so that the items where controls showed systematic change (improvement or deterioration) wouldn't need to be eliminated. This way the effects of all of these confounding factors were subtracted, so that when later investigating the change in patient performance it would be ascertained that the observed changes represent solely the MS-related change.

a) Z-score approach for standardizing scores of longitudinal change

The score standardization procedure employed in this study was adapted (with making a few changes to accommodate for the longitudinal study design) from the Multiple Sclerosis Functional Composite (MSFC) manual (Cutter et al., 1999). MSFC is a clinical outcome assessment scale widely used in everyday clinical practice (Cutter, et al., 1999). The authors of MSFC manual suggest converting raw scores on tests to z-scores (standardizing) to allow comparison of performance on tests with different units of measurement. Since the focus of this project was investigation of relative change, and not investigation of raw performance scores, therefore the raw values of performance on the test items were converted into z-scores to allow comparison of changes in performance on tests with different units of measurement

The control data was standardized by converting the raw scores of change into z-scores.

Individual Z-score = (raw score of individual's change – average raw control change) / SD of raw control change

The standardization procedure of the scores of change in patient performance was also based on the performance of the reference control group (Table 6.2.) using the same formula.

b) Establishing domain-specific change

After calculating the standardized scores of change, the next step as suggested by the creators of the MSFC, was to average the z-scores of change for each domain to allow for evaluating the domain-specific change.

Implicit in this approach was the idea that patients who deteriorate or improve on several component (domain) measures would have an overall larger change than patients who change on only one of the measures.

However, in order to allow averaging, it was necessary to establish that the tests in each domain did measure the same functions. Two tests were considered to measure the same cognitive function if they correlated. Correlations among the tests from a cognitive domain would allow averaging their scores, however, if the tests were found not to correlate, they were considered to measure separate functions. For the correlational analyses Spearman's correlation coefficient was chosen to employ. The values of patient change on each test were averaged to create standardized scores of change in five cognitive domains: verbal memory, visuospatial memory, processing speed, visuoperceptual and language.

III. Predictors of patient change in each cognitive domain

In this chapter when analysing the trajectory of longitudinal changes in performance I've expected to observe more than one trajectory of change. I hypothesized that the patients could be classified into those who improved, remained the same, and those who deteriorated in their performance on each cognitive domain. The next step was to investigate the factors that could potentially explain why some patients have changed in their performance more than others. However, the factors that potentially play a role in longitudinal changes in cognition were considered to vary and to be specific for each cognitive domain.

a) Influence of clinical predictors of patient change in each cognitive domain

In order to investigate the effect that MS-related factors have on cognition I chose to run separate multiple linear regression models for each cognitive domain. In each of the five models I've put the z-scores of patient change averaged for each domain as criterion variables, and the clinical variables as predictor variables. All values from the clinical variables were collected during the follow-up (phase II) assessment.

The regression model employed to investigate the effect of MS-related variables on cognitive change in each domain is presented:

$$\textit{Z-score of change} = \alpha + \beta_1 \textit{(MS type)} + \beta_2 \textit{(disease duration)} + \beta_3 \textit{(neurological disability)} + \beta_4 \textit{(Number of relapses in-between assessments)} + \beta_5 \textit{(depressive symptomology)} + \beta_6 \textit{(DMT uptake)} + \beta_7 \textit{(antidepressant uptake)}$$

I acknowledge that models with seven predictors and 61 observations lack power to run a multiple linear regression analysis (Green, 1991). Therefore for each domain only those predictor variables that had a relationship with the criterion variables were left in each of the models.

b) Influence of demographic and cognitive reserve predictors on performance changes in cognitive domains

Having analysed the role that MS-related variables play in predicting the average change in each of the cognitive domains, the next step was to investigate to what extent cognitive changes could be explained by factors unrelated to MS, such as demographic and cognitive reserve variables.

In order to investigate the effect that MS-unrelated factors have on cognition I again chose to run separate multiple linear regression models for each cognitive domain. In each of the five models I had put the z-scores of patient change (averaged for each domain) as criterion variables, and demographic and cognitive reserve variables as predictor variables. All values from the demographic and cognitive reserve variables were collected during the follow-up assessment, except again for the NART IQ score, which was collected at baseline.

The regression model employed to investigate the effect of MS-unrelated variables on cognitive change in each domain was proposed:

$$\textit{Z-score of change} = \alpha + \beta_1 \textit{(demographic factors)} + \beta_2 \textit{(reserve factors)}$$

The demographic factors included age and gender. The cognitive reserve factors included years of education, premorbid intelligence (NART IQ score), premorbid leisure activities score, number of languages spoken, current employment status and exercise. Therefore having included all of those variables, the full regression model of MS-unrelated predictors for change in each cognitive domain was presented:

$$\textit{Z-score of change} = \alpha + \beta_1 \textit{(age)} + \beta_2 \textit{(gender)} + \beta_3 \textit{(years of education)} + \beta_4 \textit{(premorbid intelligence)} + \beta_5 \textit{(premorbid leisure activities)} + \beta_6 \textit{(number of languages)} + \beta_7 \textit{(employment status)} + \beta_8 \textit{(exercise)}$$

As in the previous section, I acknowledge that regression models with eight predictors and 61 observations do not have sufficient power to yield reliable results (Green, 1991). Therefore for each domain only the predictor variables that correlated with the criterion variables were left in each of the models.

6.3. The trajectory and extent of cognitive change

I began the result analysis of this chapter by investigating the differences in performance between phases I and II to determine the longitudinal change. The distributions of performance scores at phase I and at phase II of control and patient participants can be seen in Appendix G. Since a large number of tests showed ceiling effect and the majority of distributions have been dramatically negatively skewed in both control and patient populations, it was chosen for the further analyses of the trajectory and extent of longitudinal change to employ non-parametric tests.

6.3.1. How much change is normal - distribution of control scores in phases I and II

As it can be seen from the Appendix G, overall the controls have exhibited similar distributions of performance scores at phases I and II. However, upon closer visual inspection it can be seen that on some tests a positive shift in the distribution was observed (SPART, SPARTD, WLGT and VOSP2), and on some other tests a negative shift in distribution was observed (SRTC, VOSP6, VOSP8, BORB5 and BC.Index).

Although this observation indicated that in the time period of 3 years the controls have improved on some tests, and deteriorated on the others, these differences in distributions of scores were only statistically significant for the VOSP6 and BC.Index items. This could imply that on those tests in general the controls have participated worse at follow-up, and that trend was evident for the whole control sample. Therefore it could be assumed that the VOSP6 and BC.Index test items are highly sensitive to cognitive ageing, or that there has been systematic bias in the administration of those tests to the whole sample.

Having analysed the overall *trends* of change in the control sample, the next step was to investigate the *extent* of individual changes in performance. The distribution of control scores in percentages in phases I and II can be seen in Table 6.1.

From analysing the changes in control performance both slight deterioration and practice effect were observed. It appears that it is normal for heathy controls to improve on some tests and deteriorate on others up to 8 per cent in this longitudinal study. This could also be part of the normal fluctuation as people do not perform exactly equally each time they are tested. On average the practice effect was most evident for the visuospatial memory, naming and processing speed tests, while the decrease related to cognitive ageing was most evident for the verbal recall, visuoperceptual and language items.

Having examined the trajectory and extent of control change, the next step was to investigate whether the same pattern of change could be observed in the patient cohort.

Table 6.1. Longitudinal change in control scores (in %) on the cognitive test items at phase I and phase II (n=23)

	Phase I	Phase II	Average
SRTL	71.12 (17.29)	71.12 (16.31)	No change
SRTC	57.08 (20.36)	57.27 (19.26)	No change
SRTD	82.61 (17.75)	77.17 (19.34)	5% decrease
SPART	69.8 (16.52)	76.38 (14.03)	7% increase
SPARTD	73.04 (18.2)	80.43 (24.58)	7% increase
SDMT	66.24 (8.46)	70.25 (8.86)	4% increase
PASAT	79.32 (21.17)	84.49 (10.92)	5% increase
WLGT	70.97 (13.82)	77.56 (16.46)	7% increase
VOSP2	77.39 (10.68)	83.04 (11)	6% increase
VOSP3	93.48 (6.29)	92.39 (7.37)	1% decrease
VOSP6	99.13 (2.45)	91.52 (5.92)	8% decrease
VOSP8	99.13 (2.88)	96.52 (5.73)	3% decrease
BORB2	91.3 (4.35)	91.16 (6.16)	No change
BORB5	92.83 (5.85)	90.54 (5.59)	2% decrease
TROG	97.02 (3.76)	96.3 (3.68)	1% decrease
BC.Index	41.67 (4.58)	34.06 (7.56)	7% decrease

Table 6.1. shows that variations in the performance on the cognitive tests were observed. The controls tend to improve on the tests of visuospatial memory, information processing speed and naming, and deteriorate on tests measuring verbal recall, visuoperceptual cognition, and spontaneous speech.

Note. For the tests where the maximum value is unknown (SDMT and WLGT), the highest value from the control and patient samples (both phases) was used to create the 100% score.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

6.3.2. Distribution of patient scores in phases I and II

As it can be seen from the Appendix G, overall the patients have exhibited similar patterns of changes in distributions of performance at phases I and II as the controls. On most of the tests the distributions at phase I and II were no different, and as in the control sample, improvement on VOSP2, and deterioration on VOSP6 was observed.

However, not only similarities, but differences could be indicated as well. While the controls showed overall improvement on both items of visuospatial memory (SPART and SPART D), the patients, however, showed only minimal improvement on SPART and a significant deterioration on SPARTD. The controls have showed a negative shift in distribution on SRTC, and BC.Index, but this was not observed in the patient sample.

Another interesting aspect was that at phase II more patients had very low scores on the PASAT, and a second peak at low values could potentially indicate that a subgroup of patients emerged who found PASAT to be very difficult at phase II, but not at phase I.

From analysing the patient performance (Table 6.2.) both improvements and deteriorations in performance were observed. On average, the patients have slightly improved on tests of verbal and visuospatial learning, and verbal recall. Most notable deteriorations in performance were observed on the test of sustained attention, and on a visuospatial test of position discrimination.

To summarize, it can be noted, that the trajectory of longitudinal change was different for the patient and control participants. The next step was to investigate to whether there were differences in the extent of change as well.

Table 6.2. Longitudinal change in patient scores (in %) on the cognitive test items at phase I and phase II (n = 82).

	Phase I	Phase II	Average change
SRTL	45.49 (22.40)	48.78 (23.21)	3% increase
SRTC	31.60 (23.03)	35.19 (23.29)	4% increase
SRTD	51.34 (25.49)	52.67 (27.63)	1% increase
SPART	58.90 (17.48)	64.23 (16.58)	5% increase
SPARTD	63.58 (21.81)	61.67 (26.01)	2% decrease
SDMT	54.88 (14.45)	53.60 (16.41)	1% decrease
PASAT	61.21 (29.35)	51.71 (35.52)	10% decrease
WLGT	59.52 (14.22)	61.72 (18.41)	2% increase
VOSP2	73.33 (13.97)	73.05 (16.52	No change
VOSP3	88.12 (10.11)	85.37 (12.32)	3% decrease
VOSP6	97.06 (5.55)	86.89 (10.57)	10% decrease
VOSP8	98.12 (91.98)	91.98 (13.64)	6% decrease
BORB2	86.67 (6.94)	86.75 (7.06)	No change
BORB5	88.48 (6.91)	86.99 (8.79)	1% decrease
TROG	94.46 (5.25)	94.38 (5.64)	No change
BC.Index	52.20 (16.14)	52.99 (12.78)	No change

Table 6.1. shows that both increases and decreases in the performance on the cognitive tests were observed. The patients tend to improve on the tests of verbal memory and visuospatial learning, and deteriorate on tests measuring sustained attention and visuospatial cognition.

Note. For the tests where the maximum value is unknown (SDMT and WLGT), the highest value from the control and patient samples (both phases) was used to create the 100% score

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

6.3.3. Comparison of the extent of the control and patient change

As it can be seen from Table 6.3., the pattern of longitudinal change in performance was not uniform, with the patients and controls exhibiting differing trajectories of change. On some tests the patients have improved in their performance more than the controls (verbal memory); on some tests the

patients have deteriorated more than the controls (spatial cognition); on several tests the controls have improved while the patients have deteriorated (processing speed); and on some tests the controls, but not the patients have deteriorated (language).

Table 6.3. Comparison of longitudinal changes in performance between control and patient participants

	Controls (n = 23)	**Patients (n = 82)**	**Difference**
SRTL	0.00 (14.42)	3.00 (20.86)	U = 888.50, p = 0.803
SRTC	0.19 (15.74)	3.23 (20.28)	U = 869.00, p = 0.686
SRTD	-5.43 (16.01)	0.94 (18.94)	U = 754.50, p = 0.185
SPART	7.65 (12.73)	5.02 (16.69)	U = 581.00, p = 0.569
SPARTD	7.39 (19.36)	-1.95 (26.51)	U = 674, p = 0.081
SDMT	**4.43 (6.71)**	**-1.31 (9.99)**	**U = 586.50, p = 0.020**
PASAT	4.70 (17.00)	-8.21 (24.94)	U = 654.50, p = 0.066
WLGT	6.59 (13.03)	2.20 (13.21)	U = 788.00, p = 0.228
VOSP2	5.65 (9.81)	0.54 (13.58)	U = 688.50, p = 0.066
VOSP3	-1.09 (7.38)	-2.50 (12.35)	U = 843.00, p = 0.535
VOSP6	-7.61 (5.81)	-10.38 (11.62)	U = 783.50, p = 0.349
VOSP8	-2.61 (6.19)	-6.25 (13.91)	U = 851.00, p = 0.543
BORB2	-0.14 (5.55)	0.00 (7.73)	U = 880.50, p = 0.820
BORB5	-2.28 (6.39)	-1.20 (8.51)	U = 750.00, p = 0.263
TROG	-0.83 (2.66)	-0.06 (0.82)	U = 710.50, p = 0.300
BC.Index	**-11.77 (9.40)**	**0.79 (19.99)**	**U = 448.00, p = 0.004**

Table 6.3. shows that on the majority of tests the patients and the controls exhibit the same trajectory and similar extent of change. Different trajectories of change were observed on tasks measuring verbal and visual recall, information processing speed and spontaneous speech. The extent of change was statistically significantly different on the SDMT and BC Index items.

Note. Comparisons were performed with the Mann Whitney U-test. Positive values indicate improvement; negative values indicate deterioration in performance

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD – items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

However, these differences reach significance only in two tests: the controls seem to show practice effect on an information processing speed task, while the patients show deterioration; and the controls show deterioration on a spontaneous speech task while the patients show no change.

On most tests the patients seemed to express similar patterns of change as the controls, attributable to practice effect, cognitive ageing, and administration differences between the phases. However, on four tests (SPARTD, SDMT, PASAT, VOSP8) the patients have deteriorated much more, and on one test (BC.Index) the patients have improved much more, with these differences being statistically significant. This could potentially indicate that MS could be associated with a more rapid decline on spatial awareness and processing speed. However, the improvement expressed by patients on the BC test could be an artefact, as due to progression of upper limb motor disability, more patients were allowed to perform it this task orally at phase II, whereas all controls have performed in writing at both phases.

In addition, I found a statistically significant difference in the distribution of BC.Index between phases I and II in the control sample, with healthy participants performing better at baseline. This could also be explained by administration differences between the researchers at phases I and II. Since there is no time limit for the BC test, different researchers may give more or less time to complete the task, or use verbal or non-verbal cues to encourage or discourage providing more text. Since the calculation of the BC.Index item is dependent on the total number of words provided, the differences in controls' BC.Index could potentially be explained by how much of text they have actually written. At phase I the 22 controls provided an average of 55.23 (SD 19.26) words, and at phase II the same people provided an average of 65.14 (SD 25.25) words, but however, this difference was not statistically significant ($t(42) = -1.464$, $p = 0.151$, 2-tailed).

6.3.4. Standardization of change in domain scores

Having standardized the change in performance for each participant on each test the next step was to establish the standardized scores of change in the cognitive domains. This was done by averaging the values of all the tests in each cognitive domain individually for each patient. However, in order to allow averaging, it was necessary to establish that the tests in each domain did measure the same

functions. Two tests were considered to measure the same function if they correlated and this would allow averaging their scores, but if they didn't correlate, they were considered to measure separate functions.

I. Correlations among the tests in each cognitive domain

In order to determine whether the items in the cognitive domains were measuring similar functions it was decided to examine the correlation matrices of the raw values on each test for each domain at both phases (Tables 6.4 – 6.8).

Table 6.4. Correlations among the test items in the verbal memory domain. Phase I and phase II patients with full data (n = 61)

		SRTL	SRTC
Phase I	SRTC	**r = 0.890, p < 0.001**	
	SRTD	**r = 0.733, p < 0.001**	**r = 0.742, p < 0.001**
Phase II	SRTC	**r = 0.992, p < 0.001**	
	SRTD	**r = 0.888, p < 0.001**	**r = 0.889, p < 0.001**

Note. Correlations adjusted for multiple comparisons (Bonferroni α level 0.017)

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval)

Table 6.5. Correlations among the test items in the visuospatial memory domain. Phase I and phase II patients with full data (n = 61)

		SPARTD
Phase I	SPART	**r = 0.753, p < 0.001**
Phase II	SPART	**r = 0.724, p < 0.001**

Abbreviations: SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD – items correct at delayed recall)

Table 6.6. Correlations among the test items in the processing speed domain. Phase I and phase II patients with full data (n = 61)

		SDMT	PASAT
Phase I	PASAT	**r = 0.577, p < 0.001**	
	WLGT	**r = 0.544, p < 0.001**	**r = 0.496, p < 0.001**
Phase II	PASAT	**r = 0.629, p < 0.001**	
	WLGT	**r = 0.577, p < 0.001**	**r = 0.443, p < 0.001**

Note. Correlations adjusted for multiple comparisons (Bonferroni α level 0.017)

Abbreviations: SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, WLGT – Category Animal Fluency task.

Table 6.7. Correlations among the test items in the visuoperceptual domain. Phase I and phase II patients with full data (n = 61)

		VOSP2	VOSP3	VOSP6	VOSP8	BORB2
	VOSP3	r = 0.347, p = 0.006				
	VOSP6	r = - 0.100, p = 0.444	r = 0.036, p = 0.784			
Phase I	VOSP8	r = 0.201, p = 0.121	r = - 0.018, p = 0.889	r = -0.186, p = 0.151		
	BORB2	r = 0.032, p = 0.808	r = - 0.103, p = 0.430	r = 0.013, p = 0.918	r = 0.090, p = 0.490	
	BORB5	r = 0.321, p = 0.012	r = 0.045, p = 0.731	r = 0.066, p = 0.614	r = 0.340, p = 0.007	**r = 0.544, p < 0.001**
	VOSP3	**r = 0.556, p< 0.001**				
	VOSP6	r = 0.167, p = 0.199	r = 0.335, p = 0.008			
Phase II	VOSP8	r = 0.363, p = 0.004	r = 0.270, p = 0.035	r = 0.141, p = 0.277		
	BORB2	r = 0.186, p = 0.151	r = 0.094, p = 0.471	**r = 0.450, p < 0.001**	r = 0.172, p = 0.184	
	BORB5	**r = 0.552, p < 0.001**	r = 0.309, p = 0.016	r = 0.361, p = 0.004	**r = 0.456, p < 0.001**	**r = 0.393, p = 0.002**

Note. Correlations adjusted for multiple comparisons (Bonferroni α level 0.003).

Abbreviations: VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task.

Table 6.8. Correlations among the test items in the language domain. Phase I and phase II patients with full data (n = 61)

		BC.Index
Phase I	TROG	r = - 0.105, p = 0.422
Phase II	TROG	**r = - 0.260, p = 0.043**

Abbreviations: TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

From analysing the relationships among the test items, I found that the tests in verbal memory, visuospatial memory, and processing speed domains were highly inter-correlated, meaning that they measure the same (or very similar) functions. However, this pattern was less visible in visuoperceptual and language domains.

The tests in the visuoperceptual domain could be considered inter-correlated, but the significance of those relationships diminished after adjusting for multiple comparisons. In addition, the lack of inter-correlations could be explained by lack of variance, as many of the visuoperceptual tests had a strong ceiling effect. Therefore it was decided that even though these tests didn't exhibit such high inter-correlations as in the memory and processing speed domains, that could be explained by the lack of variance, and thus it was concluded that the z-scores of change on the visuoperceptual domain could be averaged for further analyses.

Similarly as in the visuoperceptual domain, the tests in the language domain reached significance only in half of the correlations, and only at phase II. This could be interpreted in two ways. The first way to interpret was that at phase II the ceiling effect in TROG was slightly less evident, allowing for more variation in performance. This could explain why the two tests correlated only at phase II, and would treat phase I non-significant relationship as an artefact. The second possible interpretation was that these two tests measured different aspects of language: the TROG assessed grammar comprehension, and BC.Index item assessed spontaneous speech. It could be that these two tests were not correlated because they measured very different functions of the same domain, and therefore required different capabilities to perform them. However even though they measured different

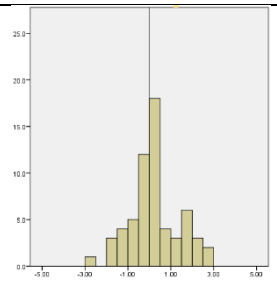
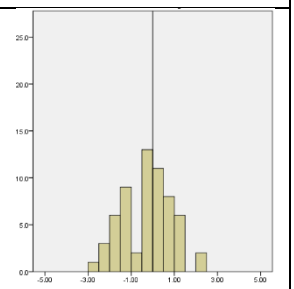
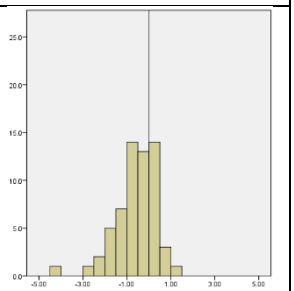
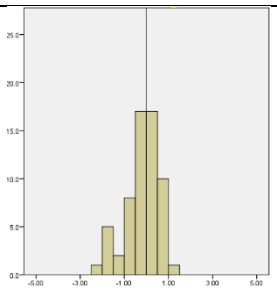
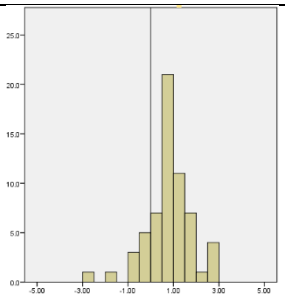
functions, expression and comprehension are both important parts of the language domain. Considering these two interpretations I chose to average the z-scores in the language domain as well, as this was also consistent with the conservative approach used throughout this project. In this instance averaging the language domain decreased the chances of picking up artefacts and making the type I error, which could potentially occur if the two items were analysed separately.

II. Calculating standardized scores of change in cognitive domains

The values of patient change on each test were averaged to create standardized scores of change in five cognitive domains: verbal memory, visuospatial memory, processing speed, visuoperceptual, and language. The averaged z-scores of domain change for each patient can be seen in Appendix G. As can be seen from Figure 6.1., on each cognitive domain both improvement and deterioration were observed.

However, on some cognitive domains the patients exhibited more deterioration than improvement (processing speed and visuoperceptual), and on other domains they exhibited more improvement than deterioration (verbal memory and language).

Figure 6.1. Distributions of criterion variables for linear regression models for patient participants (n=61). All criterion variables were in z-scores and included average change in verbal memory, visuospatial memory, processing speed, visuoperceptual and language domains

	Mean (SD)	Shapiro-Wilk test	Distribution
Verbal memory	0.218 (1.131)	W = 0.979, p = 0.394	
Visuospatial memory	-0.276 (1.094)	W = 0.980, p = 0.429	
Processing speed	-0.542 (0.933)	W = 0.945, p = 0.008	
Visuoperceptual	-0.196 (0.754)	W = 0.944, p = 0.007	
Language	0.778 (0.940)	W = 0.927, p = 0.001	

Note. The vertical line represents no change.

6.4. Predictors of patient change in each cognitive domain

In section 6.2. of this chapter I've established that the patients have changed in their performance in more than one trajectory – there were patients who improved, remained the same, and those who deteriorated in their performance in each cognitive domain. The next step was to investigate the factors that could potentially explain why some patients have changed in their performance more than others. However, those factors that play a role in cognitive changes can vary, and can also be specific for each cognitive domain.

6.4.1. MS-related predictors

In order to investigate the effect that MS-related factors have on cognition I chose to run separate multiple linear regression models for each cognitive domain. In each of the five models I've put z-scores of patient change averaged for each domain as criterion variables, and clinical variables as predictor variables. The averaged z-scores of domain change were deemed suitable to be criterion variables in a linear regression model, as they had a wide range and appeared to be sufficiently normally distributed after visual inspection (Figure 6.1.)

I. Influence of time interval in-between assessments on clinical variables

Before continuing with any further analyses it was necessary to determine whether any links could be identified between the MS-related variables and the time intervals in-between assessments. Namely, whether longer duration in-between the phase I and phase II assessments correlated with any of the clinical variables. As it can be seen from Table 6.4., none of the clinical variables were linked to the time interval in-between assessments.

Table 6.9. Correlations between clinical predictor variables and the time interval in-between assessments for patient participants (n=61)

	Interval in-between assessments (years)
Predictor variables	
MS type	r = 0.097, p = 0.348
MS duration	r = - 0.034, p = 0.715
EDSS	r = - 0.034, p = 0.728
Number of relapses	r = - 0.035, p = 0.738
BDI-II	r = 0.102, p = 0.265
DMT uptake	r = 0.163, p = 0.135
Antidepressant uptake	r = 0.070, p = 0.526

Table 6.4. indicates that none of the clinical variables were linked to the time interval in-between assessments

Note. For all analyses Kendall Tau correlations were employed

Abbreviations: EDSS – Expanded Disability Status Scale, BDI-II – Beck's Depression Inventory 2nd Ed., DMT – Disease modifying treatment

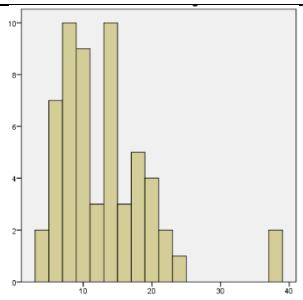
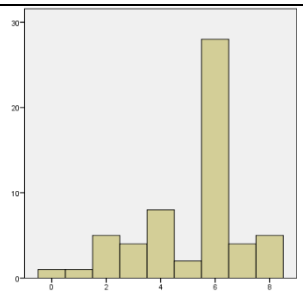
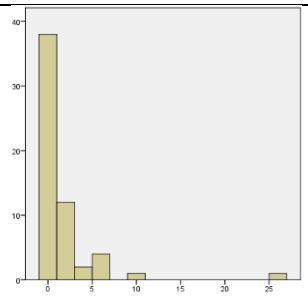
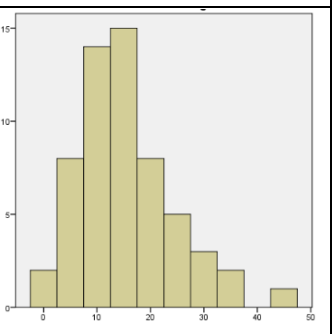
II. Preliminary investigation of relationships between potential predictor and criterion variables

For the following analyses of linear relationships between predictor and criterion variables the non-parametric tests were chosen to employ for all numeric variables (MS duration, EDSS, number of relapses, BDI-II score), as the assumption of bivariate normality could be justified only for criterion, but not predictor variables in this sample. The distributions of MS-related variables can be seen in Figure 6.2.

Kendall's Tau correlation was chosen to employ since it has been shown to be less sensitive to error and to work better with smaller sample sizes than other measures of linear relationships (Bonett & Wright, 2000). For categorical variables I used comparisons of mean (ANOVA for MS type, and Mann Whitney U-test for DMT and antidepressant uptake) to establish associations between predictor and criterion variables.

Figure 6.2. Distributions of clinical predictor variables for patient participants (n=61)

a) Distributions of numeric predictor variables

Variable	Mean (SD)	Shapiro-Wilk test	Distribution
Disease duration (years)	12.59 (6.91)	W = 0.849, p < 0.001	
EDSS	5.19 (1.86)	W = 0.945, p = 0.01	
Number of relapses	1.33 (3.68)	W = 0.877, p < 0.001	
BDI – II	15.12 (9.38)	W = 0.394, p < 0.001	

b) Frequencies of categorical predictor variables

Variable	Frequency
MS type	16 (26%) PPMS, 18 (30%) SPMS, 27 (44%) RRMS
DMT	27 (44%) no, 34 (56%) yes
Antidepressants	39 (64%) yes, 22 (36%) no

Abbreviations: EDSS – Expanded Disability Status Scale, BDI-II – Beck's Depression Inventory 2nd Ed., DMT – Disease modifying treatment

As it can be seen from Table 6.5., the changes in each cognitive domain were associated with different clinical variables. The change in verbal memory domain was linked to the levels of neurological disability and depressiveness. The change in visuospatial memory didn't seem to be linked to any of the MS-related variables. The change in processing speed was linked to the MS type only. The change in visuospatial domain was linked to neurological disability, MS type and DMT. The change in the language domain was linked to the level of depressiveness.

Table 6.10. Relationships between clinical predictor and criterion (average change in a domain) variables for patient participants

	Criterion variables				
	Verbal memory	**Visuospatial memory**	**Processing speed**	**Visuoperceptual**	**Language**
Predictor variables					
MS duration	r = - 0.123, p = 0.177	r = 0.164, p = 0.072	r = - 0.140, p = 0.123	r = - 0.001, p = 0.315	r = - 0.069, p = 0.446
EDSS	**r = - 0.227, p = 0.018**	r = - 0.047, p = 0.629	r = - 0.078, p = 0.419	**r = - 0.226, p = 0.019**	r = 0.103, p = 0.283
Number of relapses	r = - 0.163, p = 0.108	r = - 0.074, p = 0.466	r = 0.077, p = 0.447	r = 0.107, p = 0.293	r = - 0.019, p = 0.848
BDI-II	**r = -0.304, p = 0.001**	r = - 0.071, p = 0.428	r = - 0.094, p = 0.295	r = 0.068, p = 0.447	**r = - 0.181, p = 0.042**
MS type	F = 2.594, p = 0.083	F = 2.258, p = 0.114	**F = 5.184, p = 0.008**	**F = 14.264, p < 0.001**	F = 1.844, p = 0.167
DMT uptake	U = 394, p = 0.345	U = 372, p = 0.206	U = 411, p = 0.486	**U = 286, p = 0.012**	U = 421, p = 0.581
Antidepressant uptake	U = 293, p = 0.074	U = 346, p = 0.339	U = 329, p = 0.221	U = 364, p = 0.500	U = 240, p = 0.294

Table 6.5. shows that the changes in each cognitive domain were associated with different clinical variables. The change in verbal memory domain was linked to the levels of neurological disability and depressiveness. The change in visuospatial memory didn't seem to be linked to any clinical variables. The change in processing speed was linked to MS type only. The change in visuoperceptual domain was linked to neurological disability, MS type and DMT. The change in the language domain was linked to the level of depressiveness.

Note. Kendall Tau correlations for numeric variables, ANOVA for MS type, and Mann-Whitney U-test for medication uptake comparisons were employed

Abbreviations: EDSS – Expanded Disability Status Scale, BDI-II – Beck's Depression Inventory 2nd Ed., DMT – Disease modifying treatment

Based on these preliminary investigations of the relationships between predictor and criterion variables, different criterion variables were selected to enter into each multiple linear regression model.

6.4.2. Verbal memory

Based on the presented associations (Table 6.5.), the following model for the changes in verbal memory domain was proposed:

$$Z\text{-score of change in verbal memory} = \alpha + \beta_1 (EDSS) + \beta_2 (BDI\text{-}II)$$

Before running this model the predictor variables were checked for multicollinearity. EDSS and BDI-II were not correlated (r = 0.147, p = 0.134, Kendall Tau, 2-tailed). Durbin-Watson test values indicated that there was no autocorrelation (d = 1.725), and it was concluded that it was safe to proceed with the multiple linear regression model (Durbin & Watson, 1950).

When average change in verbal memory was predicted it was found that both EDSS score (Beta = -0.278, p = 0.023) and BDI-II score (Beta = - 0.317, p = 0.010) were significant predictors. The overall model fit was $R^2_{adj.} = 0.212$, p = 0.001. As it can be seen from Figure 6.3, deterioration in performance on verbal memory tests can be linked to higher neurological disability and higher levels of depression.

Figure 6.3. Linear relationship between the effects of neurological disability and depressiveness (at phase II) on the change in performance on the verbal memory domain

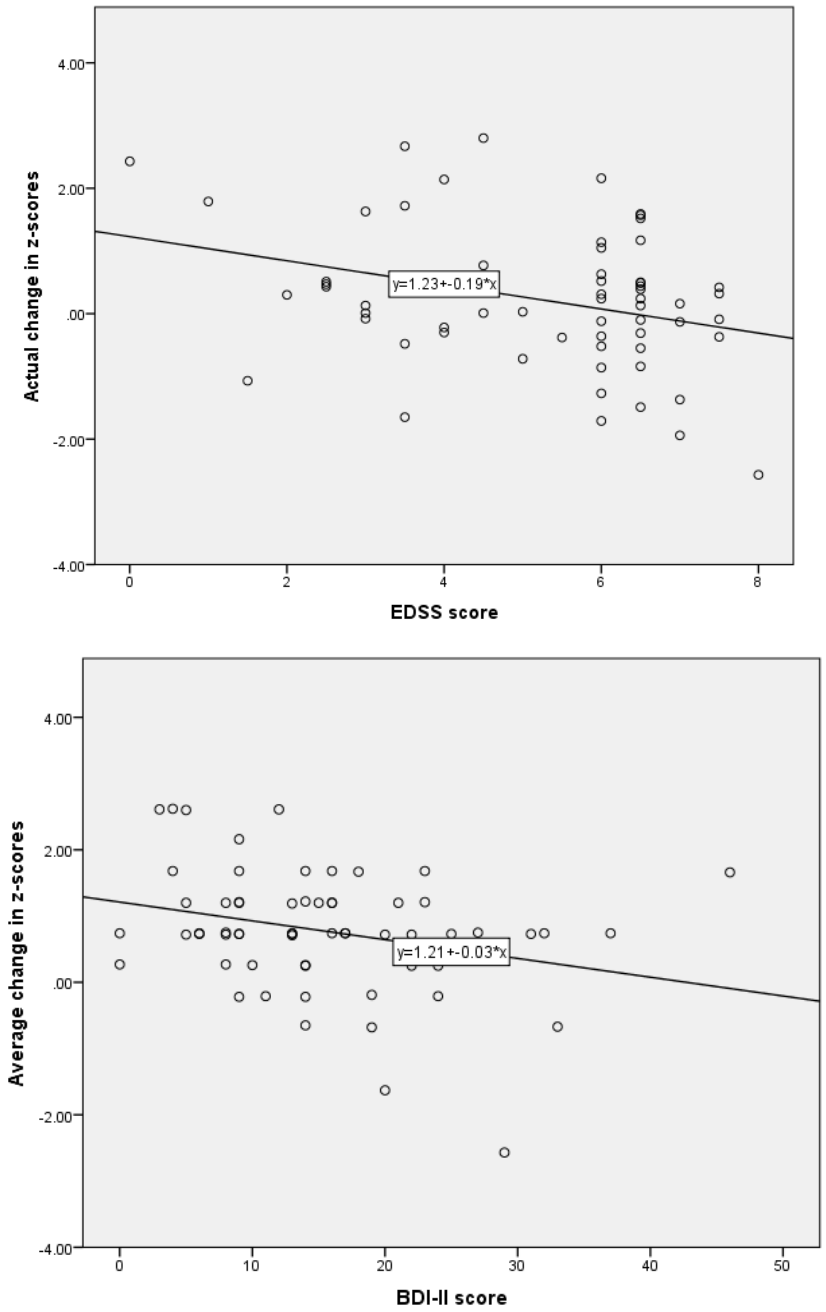


Figure 6.3. shows that higher EDSS and BDI-II scores predict negative change in performance on verbal memory tests.

Abbreviations: EDSS – Expanded Disability Status Scale, BDI-II – Beck's Depression Inventory, 2nd Ed.

6.4.3. Visuospatial memory

Since the average change in the visuospatial memory domain was not shown to be linked to any of MS related variables (Table 6.5.), there was not enough justification to support running a linear regression model.

6.4.4. Processing speed

Based on the associations presented in Table 6.5., the following model for the longitudinal changes in processing speed was proposed:

$$Z\text{-score of change in processing speed} = \alpha + \beta_1 (MS\ type)$$

In order to include the 3-level MS type variable (RRMS, PPMS, SPMS) into the linear regression model, dummy coding was used, and RRMS was used as a baseline category to which PPMS and SPMS were compared. Durbin-Watson test values indicated that there was no autocorrelation (d = 1.744) thus it was considered suitable to proceed with the linear regression model.

When the average change in processing speed was predicted it was found that progressive MS types had a negative effect. This indicated that patients with progressive types of MS suffer from poorer outcome on the processing speed tests. However, PPMS (Beta = - 0.276, p = 0.039) had a smaller effect than SPMS (Beta = - 0.813, p = 0.003). The overall model fit was $R^2_{adj.} = 0.152$, p = 0.008, and a visual representation of the model can be seen in Figure 6.4.

Figure 6.4. Linear relationship between MS type and change in processing speed domain

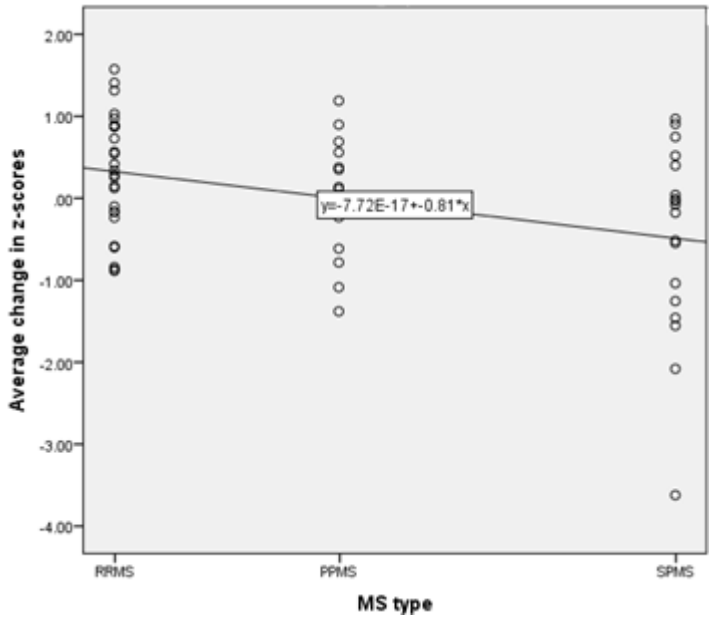


Figure 6.4. shows that progressive MS types were associated with decrease in processing speed

Abbreviations: RRMS – Relapsing-remitting MS, PPMS – Primary progressive MS, SPMS – Secondary progressive MS

6.4.3. Visuoperceptual domain

The associations presented in Table 6.5. justified running the following linear regression model for predicting average change in the visuoperceptual domain:

$$\textit{Z-score of change in visuoperceptual domain} = \alpha + \beta_1 \textit{(MS type)} + \beta_2 \textit{(DMT uptake)}$$

In this model again dummy coding was used for the MS type variable as it had three levels (RRMS, PPMS and SPMS), and RRMS was used as a baseline category. Fort the DMT uptake variable there were two categories (not taking and taking DMT drugs), and not taking DMT drugs was used as a baseline category. Durbin-Watson test values indicated that there was no autocorrelation (d = 1.702).

The overall model fit was $R^2_{adj} = 0.374$, $p < 0.001$, and a visual representation of this model can be seen in Figure 6.5.

Figure 6.5. Linear relationship between MS type and DMT uptake as predictor variables for change in visuoperceptual domain

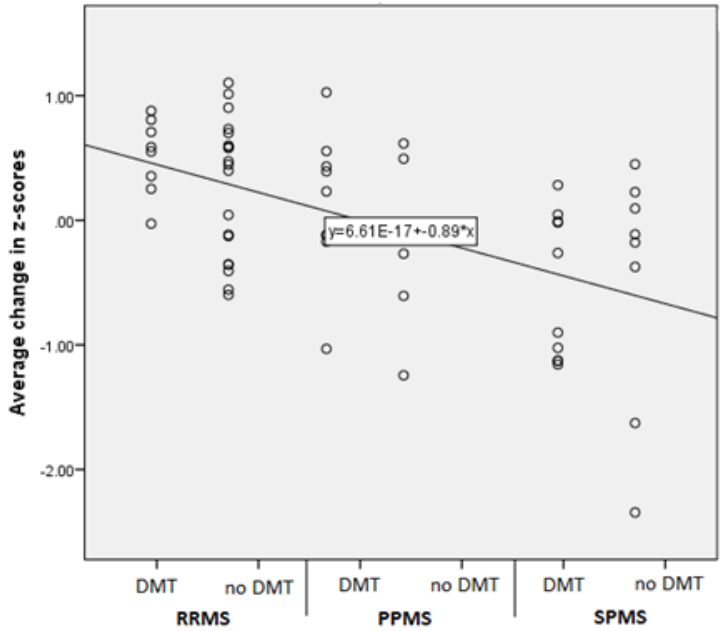


Figure 6.5. shows that SPMS type was associated with decrease in performance, but using DMT was associated with better performance on visuoperceptual tests

Abbreviations: RRMS – Relapsing-remitting MS, PPMS – Primary progressive MS, SPMS – Secondary progressive MS, DMT – disease modifying treatment

When the average change in the visuoperceptual domain was predicted, it appeared that progressive MS type had a negative effect, while adhering to DMT had a positive effect. Comparing to RRMS, SPMS was predictive of deterioration in the visuoperceptual domain (Beta = - 0.544, $p < 0.001$), while PPMS appeared to have very limited effect (Beta = - 0.013, $p = 0.908$). In addition, it was shown that taking DMT drugs had a positive effect, and could be associated with improvement in performance on visuoperceptual tests (Beta = 0.327, $p = 0.05$), although this effect was borderline significant.

Adherence to DMT and MS type were not related (X^2 (2) = 4.206, $p = 0.122$), although a higher percentage of RRMS patients were on DMT (43.8% PPMS, 44.4% SPMS, 70.4% RRMS).

6.4.6. Language

Based on the findings in Table 6.5., it was decided to include only the level of depressiveness as a sole predictor variable into the regression model predicting the average longitudinal change in the language domain:

$$Z\text{-score of change in language domain} = \alpha + \beta_1 (\text{BDI-II score})$$

For this model the Durbin-Watson test value indicated that there was no autocorrelation (d = 2.085).

When the average change in the language domain was predicted, BDI-II score had a statistically significant effect (Beta = - 0.278, p = 0.030), indicating that higher BDI-II scores were predictive of decrease in performance on the language tests. The overall model fit was $R^2_{adj} = 0.077$, p = 0.030, and a visual representation of the model can be seen in Figure 6.6.

Figure 6.6. Linear relationship between levels of depressiveness as predictor variable for change in language domain

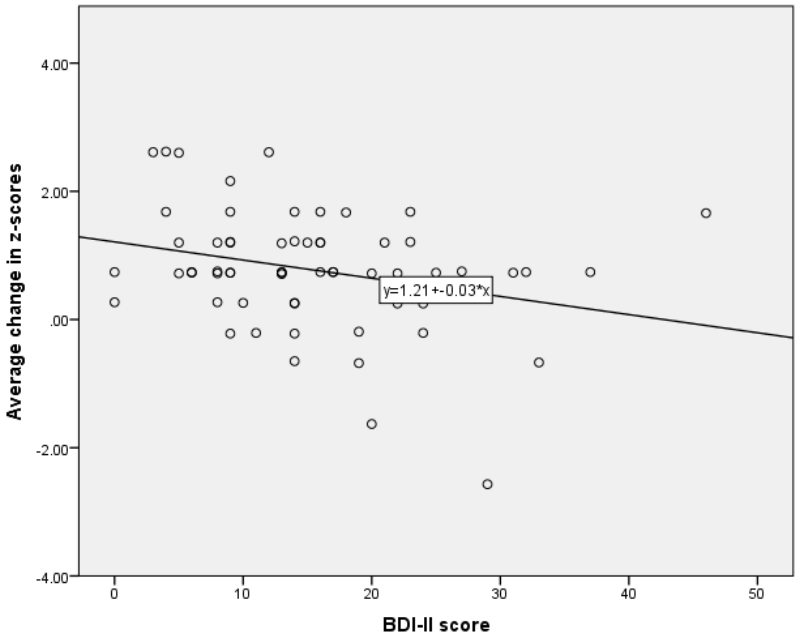


Figure 6.6. shows that higher BDI-II scores were predictive of decrease in performance on language tests.

Abbreviations: BDI-II – Beck's Depression Inventory, 2nd Ed.

6.4.7. MS-related predictors of domain-specific cognitive change. Summary

In this section I've identified that changes in performance on verbal memory, processing speed, visuoperceptual and language domains could be to some extent predicted by different MS-related variables, but, however, the change in performance on tests from visuospatial memory domain could not.

In the verbal memory domain higher EDSS and BDI-II scores predicted deterioration in performance. This indicated that in our sample neurological disability and level of depressiveness were predictive of poorer performance on the verbal memory tests.

From analysing the regression model with the change in processing speed performance as the criterion variable, it was found that progressive MS types were associated with decrease in processing speed. In the span of 3 years both PPMS and SPMS patients tend to decrease in their performance on tests of processing speed more than RRMS patients.

In the visuoperceptual domain it appeared that SPMS type and DMT uptake were associated with decrease in performance on visuoperceptual tests. Here the SPMS type was statistically significantly more linked to decrease in performance on visuoperceptual tests that other MS subtypes. This analysis had also indicated that patients who use DMT tend to have a better outcome on the visuoperceptual tests than those who don't take DMT drugs. However, this finding was borderline significant and therefore should be proceeded with caution.

In the language domain higher BDI-II scores were predictive of decrease in performance on the language tests. This indicated that individuals who were more depressed tend to deteriorate in their performance on the language tests.

To conclude, clinical variables do have an effect on cognition, but different variables affect different cognitive domains. This could potentially explain confounding findings from previous studies that related changes in cognition to different clinical factors (Table 1.1.). The discrepancy of findings linking cognitive deterioration to different clinical and non-clinical variables in previous research

could be potentially explained by which tests those studies have employed, i.e. which cognitive domains were assessed.

6.5. Demographic and cognitive reserve predictors of patient changes in each cognitive domain

After analysing the effect that the MS-related variables have on longitudinal changes in performance on each of the cognitive domains, the next step was to investigate whether the demographic and cognitive reserve variables help explain some of the changes as well. In this section I aimed to explore whether age, gender, premorbid IQ, premorbid leisure activities, years of education, exercise, employment status and the number of languages spoken could help predict the longitudinal changes.

I. Correlations between predictor and criterion variables

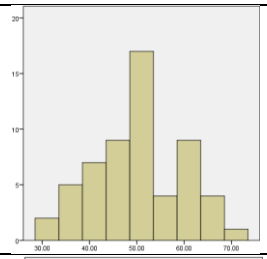
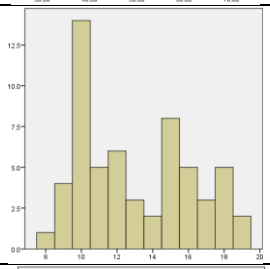
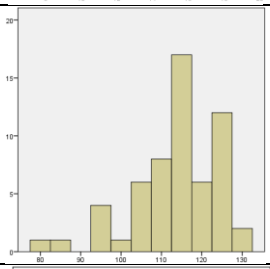
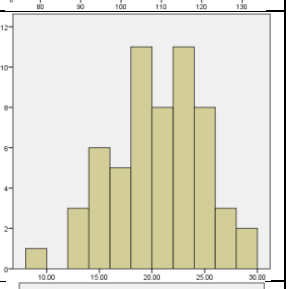
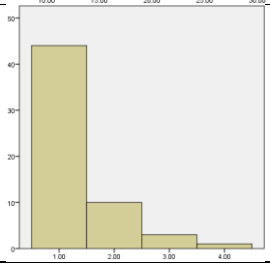
As before, I acknowledge that models with eight predictors and 61 observations did not have sufficient power to run a multiple linear regression analysis (Green, 1991). Therefore for each domain only the predictor variables that had linear relationship with the criterion variables were left in each of the models.

For these analyses Kendall's Tau correlation test was chosen to employ for all numeric variables (age, years of education, NART IQ score, premorbid leisure activities score, and number of languages), as the assumption of bivariate normality could not be justified for predictor variables in this sample (Figure 6.7.). For categorical variables (gender, employment status and exercise) I used non-parametric comparisons of mean (Mann-Whitney U-test).

As it can be seen from Table 6.6., the changes in performance on four cognitive domains (verbal memory, visuospatial memory, processing speed and language) weren't associated with demographic and cognitive reserve variables. A potential link between change in visuoperceptual domain, age and employment was identified, and these two predictors were used to include into the model. The change in visuoperceptual domain model was the only one considered suitable for further investigation.

Figure 6.7. Distributions of demographic and cognitive reserve predictor variables for linear regression models for patient participants (n=61)

a) Distributions of numeric variables

Variable	Mean (SD)	Shapiro-Wilk test	Distribution
Age (years)	49.98 (9.32)	W = 0.983, p = 0.601	
Years of education	13.05 (3.15)	W = 0.916, p = 0.001	
NART premorbid IQ	113.53 (10.20)	W = 0.928, p = 0.002	
Premorbid leisure	20.03 (4.36)	W = 0.987, p = 0.804	
Number of languages spoken	1.33 (0.66)	W = 0.559, p < 0.001	

b) Frequencies of category variables

Variable	Frequency
Gender	32 (52%) female, 29 (48%) male
Exercise	33 (54%) no, 28 (46%) yes
Employment	27 (44%) no, 34 (56%) yes

Table 6.11. Correlations between predictor (demographic and reserve) and criterion (average change in a domain) variables for patient participants (n=61)

	Criterion variables				
	Verbal memory	**Visuospatial memory**	**Processing speed**	**Visuoperceptual**	**Language**
Predictor variables					
Demographic					
Age	r = - 0.063, p = 0.417	r = - 0.103, p = 0.198	r = - 0.138, p = 0.079	**r = - 0.169, p = 0.033**	r = - 0.004, p = 0.963
Gender	U = 371, p = 0.179	U = 419, p = 0.516	U = 464, p = 0.999	U = 379, p = 0.220	U = 445, p = 0.784
Reserve					
Years of education	r = 0.018, p = 0.828	r = 0.018, p = 0.829	r = 0.038, p = 0.651	r = 0.137, p = 0.103	r = 0.045, p = 0.614
NART premorbid IQ	r = - 0.096, p = 0.218	r = - 0.006, p = 0.941	r = - 0.044, p = 0.574	r = 0.110, p = 0.168	r = 0.084, p = 0.319
Premorbid leisure activities	r = - 0.019, p = 0.832	r = 0.015, p = 0.871	r = 0.091, p = 0.312	r = 0.017, p = 0.851	r = 0.055, p = 0.540
Number of languages	r = - 0.175, p = 0.091	r = 0.038, p = 0.715	r = 0.060, p = 0.561	r = - 0.036, p = 0.728	r = 0.033, p = 0.753
Employment status	U = 354, p = 0.260	U = 369.5, p = 0.371	U = 351, p = 0.241	**U = 148, p < 0.001**	U = 393, p = 0.589
Exercise	U = 352, p = 0.111	U = 381, p = 0.241	U = 365, p = 0.160	U = 451, p = 0.874	U = 377, p = 0.219

Table 6.6. indicates that in general the change in performance on cognitive tests wasn't associated with demographic and reserve variables. A link between change in visuoperceptual domain, age and employment needed to be further investigated.

Note. Kendall Tau correlations for numeric variables, Mann-Whitney U-test for binary variables.

6.5.1. Visuoperceptual domain

Based on the presented associations (Table 6.6.) the following model for the changes in visuoperceptual domain was proposed:

$$Z\text{-score of change in verbal memory} = \alpha + \beta_1 (age) + \beta_2 (employment\ status)$$

For the employment status variable there were two categories (not employed and currently employed), and not employed was used as a baseline category.

Before running this model the predictor variables were checked for multicollinearity. Age and employment status were found to be negatively correlated (r = - 0.239, p = 0.027, Kendall Tau, 2-tailed). However, it was considered to assume that older people in our sample have retired due to older age and lost their employment this way. But after checking that assumption I found that out of the 39 (of 61) unemployed patients only four were above the retirement age of 65 years. Therefore it was concluded that the unemployment in this sample represented retirement due to MS, and not due to reaching retirement age. Durbin-Watson test values indicated that there was no autocorrelation (d = 1.725).

When average change in visuoperceptual domain was predicted it was found that being employed was the only significant predictor (Beta = 0.812, p = <0.001), and that age had no significant effect (Beta = 0.001, p = 0.946). The overall model fit was $R^2_{adj.} = 0.244$, $p < 0.001$. As it can be seen from Figure 6.8, deterioration in performance on visuoperceptual tests was linked to being unemployed.

Figure 6.8. Linear relationship between employment status and change in visuoperceptual domain

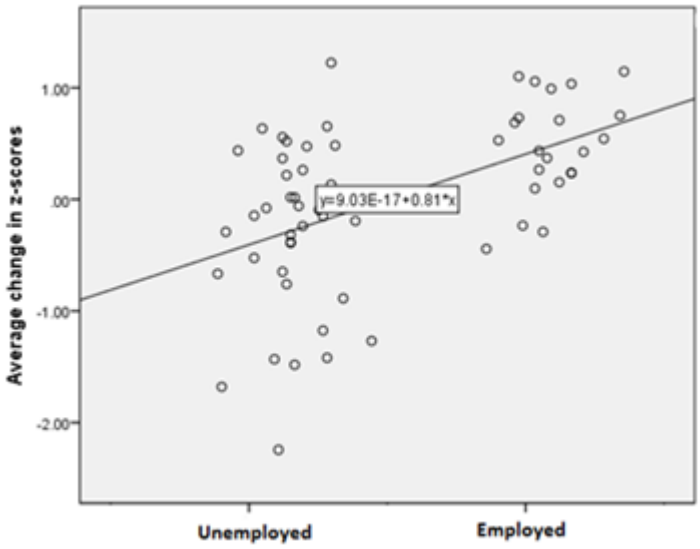


Figure 6.8. indicates that being unemployed was shown to be the sole predictor of deterioration visuoperceptual domain

This analysis indicated that impairments in visuoperceptual domain could cause participants problems at work and eventually lead to unemployment.

6.5.2. Demographic and cognitive reserve predictors of domain-specific change. Summary

In this section I've identified that changes in performance on verbal memory, visuospatial memory, processing speed and language domains couldn't have been predicted by demographic and cognitive reserve factors. The change in performance on tests from visuoperceptual domain could be predicted by employment status, but however, the directionality of this relationship was unclear. It could either be that continuing to be employed stimulated cognitive abilities required for performing visuoperceptual tests, but it could also be that failure to perform functions assessed by visuoperceptual tests resulted in failure in performing work-related tasks (such as reading or identifying numbers) therefore resulting in patients losing their jobs.

Besides this, it can be concluded that cognitive reserve factors didn't have an effect on longitudinal change in cognitive performance. It may be that cognitive reserve factors have an effect on the overall ability, but not on the rate of change – if an impairment in cognitive function develops, it progresses at the same rate for everyone, high or low premorbid ability.

6.6. Chapter summary

In this chapter I've identified the trajectory, extent and predictors of cognitive change in MS. To begin with, I found is that there was no clear pattern of what trajectory and extent the change takes, and from what can be inferred from our sample, there was also no single predictor for cognitive change, as the impairments on different domains were linked to different MS-related variables.

From analysing the trajectories of change I've shown that both pwMS and healthy controls exhibited changes in performance on the cognitive tests. Both controls and pwMS showed increased and decreased performance, and the extent of the changes differed depending on a cognitive test. In this chapter I've produced a standardized way to estimate the amount of change, and this method was shown to allow comparisons of longitudinal change across the cognitive domains. By standardizing the scores with reference to control performance I have accounted for variation that could have been caused by cognitive ageing, practice effect, and differences in test administration.

From analysing the predictors of cognitive change I've identified that the change on each cognitive domain could be predicted by different MS-related variables. The change in *verbal memory domain* was linked to the levels of *neurological disability and depressiveness*. The change in *visuospatial memory* didn't seem to be linked to any clinical variables. The change in *processing speed* was linked to *MS type* only (SPMS, and to lesser extent PPMS). The change in *visuoperceptual domain* was linked to *neurological disability, MS type and DMT*. The change in the *language domain* was linked to the levels of *depressiveness*.

In contrast to the effect of clinical variables, the demographic and cognitive reserve variables did not have predictive value on the cognitive change. The only exception was the change in visuoperceptual domain, where poorer performance was found to be associated with unemployment. However, the directionality of this relationship was unclear.

Chapter Seven. Self-perception of cognitive change in pwMS

7.1. Chapter overview

From analysing the longitudinal change in Chapter Six it was found that both improvements and deteriorations in performance were observed in this patient sample, and that they had been caused by multiple factors. The subsequent step was to investigate whether the patients themselves were aware of the trajectory and extent of their cognitive change, and the factors that have had an effect on those estimations.

The first attempt was to replicate the findings from the previous studies on insight where the patients had been assessed solely with the BRBN battery (composed of tests assessing verbal memory, visuospatial memory, processing speed, attention and verbal fluency). Then the findings on the self-perception of cognitive difficulties in performance on the BRBN battery were compared to the tests which haven't been studied before in this context in MS, namely the visuoperceptual and language tests.

Since it is common for healthy adults to exhibit some level of inaccuracies in their estimations, the patient self-estimates of longitudinal change in cognition were compared not only to their actual performance, but also to that exhibited by the reference control population.

7.2. Methods

7.2.1. Participants

For all of the analyses in this chapter only the participants with full data on all cognitive assessments were chosen to employ. This way it was controlled that all participants had gone through the exactly same tests and therefore provided their cognitive performance self-estimates for the same assessments, allowing for valid between-domain comparisons.

For the analyses comparing the actual change with the perceived change, the scores from 61 patient and 15 controls were used. For further analyses of the predictors of self-perceptions of longitudinal change, only the demographic and disease data from the patient participants was used.

7.2.2. Data format

In this chapter only the self-perceptions of *longitudinal change* in performance have been analysed. In order to address the doubts regarding the reliability of collecting the self-awareness data for phase I *at phase II*, we have further investigated the patient and control abilities to estimate their performance at phase II in the supplementary analysis (Appendix I). However, since no major discrepancies have been observed between the patients' and controls' abilities to estimate their cognitive performances, it was considered safe to continue with the analyses of self-perceptions of longitudinal change.

I. Actual change in performance

The actual change in performance on the three groups of tests (the BRBN battery, visuoperceptual and language) was analysed as z-scores as had been done in Chapter Six.

The standardized values of change on the language and visuoperceptual tests for the patient and control participants were employed from Chapter Six, section 6.3.4. and can be seen in Appendix H. The standardized values of change on the BRBN battery were calculated separately by averaging the standardized values of change in performance on its component test items: SRTL, SRTC, SRTD, SPART, SPARTD, SDMT, PASAT, and WLGT. In the previous chapters these test items were used to produce the domain scores for verbal memory, visuospatial memory, and processing speed.

These cognitive tests comprising the BRBN battery are widely employed in MS research and impairments on the functions assessed by them are considered to represent a classical view on MS-related cognitive impairment. Even though the BRBN battery includes tests that assess multiple cognitive domains (verbal memory, visuospatial memory, processing speed, attention, and category naming), it was decided to study patient self-perception of cognitive change on this battery as a whole, to allow comparison with previous work on insight in MS. The previous studies (such as Christodoulou et al., 2005; Sherman, Rapport, & Ryan, 2008; Julian, Merluzzi, & Mohr, 2007;

Carone, Benedict, Fishman, & Weinstock-Guttman, 2005) have all used solely the BRBN battery to assess cognitive deficits in MS, and collect data on awareness of these deficits. The estimates of self-awareness of longitudinal change for the visuoperceptual and language domains were collected separately, as the insight into the deficits on these cognitive domains have never been studied before in MS.

II. *Perceived change in performance*

The estimates of perceived change in performance were collected on three groups of cognitive tests: the BRBN battery, and the visuoperceptual and language tests. Two types of estimates were collected – the self-estimation of perceived levels of performance at phase I, and the self-estimation of perceived levels of performance at phase II. The participants had provided both of these measures at phase II, therefore the self-estimates of performance at phase I were collected retrospectively.

The self-estimates of longitudinal change in cognition were calculated using the following formula:

Self-estimate of longitudinal change = Self-estimate of perceived performance at phase II – Self-estimate of perceived performance at phase I

By calculating the self-estimates of perceived longitudinal change in cognition this way, the positive differences indicated perceived improvement, and the negative differences indicated perceived deterioration in performance. Since the range for both estimations (at phase I and at phase II) were on a scale from 0 to 100, theoretically the values of self-estimations of longitudinal change would range from [– 100] to [+ 100], and the further they were from 0, the larger the perceived change.

7.2.3. Group differences in actual change in performance

First the standardized scores of the *amount of actual change* between the patient and control participants were compared with the aim of investigating any differences that could be found b*etween the patients and controls* in how much they have actually changed on the three groups of cognitive tests. All comparisons were performed with the Mann-Whitney U-test. The dependent variables were the changes in performance on the BRBN tests, the visuoperceptual tests and the language tests; and the independent variable was being a patient or a control.

As a next step the amount of the *actual change in performance* on the three groups of tests was also compared for the patient and control samples separately. I wanted to know whether there have been differences in the amount of change *between the three groups of tests*. Namely, this way I've tested whether the participants have changed on some of the groups of tests more than on others. Those analyses were performed with the Related-Samples Friedman's analysis of variance by ranks twice, first for the patient participants, then separately for the control participants.

7.2.4. Group differences in the perceived changes in performance

After estimating the amounts of the actual change, I wanted to know whether there have been differences between the self-estimates of perceived change. The result analysis was conducted in the same manner, but this time the differences in self-estimates of change instead of the actual change were analysed.

First I've compared the *self-estimates of change between the patient and control participants*. I wanted to know whether there were any differences between the patients and the controls in how much they have thought that they have changed on the three groups of tests. As has been done with comparisons of actual change, the comparisons of perceived change were performed with the Mann-Whitney U-test. The dependent variables were the perceived changes in performance on the BRBN tests, the visuoperceptual and the language tests; and the independent variable was being a patient or a control.

The perceived change on the three groups of tests was also compared for the patient and control samples separately. I wanted to know whether there have been differences in the amount of perceived change *between the three groups of tests.* These analyses were performed with the Related-Samples Friedman's analysis of variance by ranks twice, first for the patient participants, then separately for the control participants.

7.2.5. Relationship between the actual and the perceived change in performance

Having analysed the actual changes and the perceived changes in performance separately, the next step was to investigate the relationship between the self-estimated and the actual longitudinal changes

in cognition. This was done separately for each group of cognitive tests in both patients and controls, first by visually inspecting the trajectories of the actual longitudinal change and the perceived change, and then by running multiple correlations between those measures. Kendall Tau correlation was employed due to small sample sizes and non-normal distribution in the control cohort.

7.2.6. Predictors of perceived change

Previous studies have identified that about 60% of pwMS believe that they have developed cognitive difficulties. I wanted to know what makes some patients believe that their performance on the cognitive tests has changed. The next step was to investigate the factors considered to be related to perceiving changes in cognitive abilities in the patient population.

The following multiple linear regression model for the patient participants was devised:

$$\textit{Perception of longitudinal change in patients} = \alpha + \beta_1 \textit{(executive functioning)} + \beta_2 \textit{(depressive symptomology)} + \beta_3 \textit{(neurological disability)} + \beta_4 \textit{(Number of relapses in-between assessments)} + \beta_5 \textit{(MS impact)}$$

In this model it was predicted that poor executive functioning would be associated with underestimating the actual change. However, higher scores on the depression, MS impact, neurological disability scales, and more relapses in-between assessments, would be associated with perceived decline in cognitive functioning in all groups of tests. I anticipated these factors to have similar predictive values in all three models, on all three groups of cognitive tests.

In this model the executive functioning was measured by the WLGT test, the depressiveness was measured with the BDI-II scale, and the neurological disability was measured with the EDSS scale. MS impact was measured with the MSIS-29 scale, and the number of relapses in-between the assessments was noted for the RRMS and SPMS participants (n=45).

I do acknowledge that the WLGT is a category naming task that is considered to be a test of executive functioning in some of the studies, but not on the others. However, since no other better suitable test of executive functioning had been included into the phase I battery, it was decided to proceed with WLGT as an indicator of executive functioning in this model.

7.3. Results

In this chapter the z-scores of change on the individual BRBN battery, visuoperceptual and language tests were averaged to produce the values of actual change for each groups of tests for both patient and control participants. As it can be seen from the data presented in Appendix H, both patient and control participants weren't exactly accurate in their estimations of their personal performance.

7.3.1. Actual changes in performance

I. Comparison of the change in performance scores between the patients and controls

First I've compared the standardized scores of the actual change between the patient and control participants. I wanted to know whether there have been differences between how much the patients and the controls have actually changed on the groups of cognitive tests.

As it can be seen from Figure 7.1., both controls and patients have exhibited improvements and deteriorations in their performances on all groups of tests. No differences between the patients and the controls were found in the patterns of longitudinal change in their performance on the BRBN and visuoperceptual tests.

The main difference was on the change in performance on the language tests, where more controls than patients have deteriorated. The difference in patient and control distributions on the change in performance on the language tests was statistically significant (Kolmogorov-Smirnov $Z = 1.744$, $p = 0.005$). This needs to be taken into account when analysing the predictors of change in language performance, as the patients could be considered unchanged in their performance in comparison to the reference control group, who for some reason have deteriorated.

Figure 7.1. Comparison of standardized scores of actual change in performance between control and patient participants for each group of cognitive tests

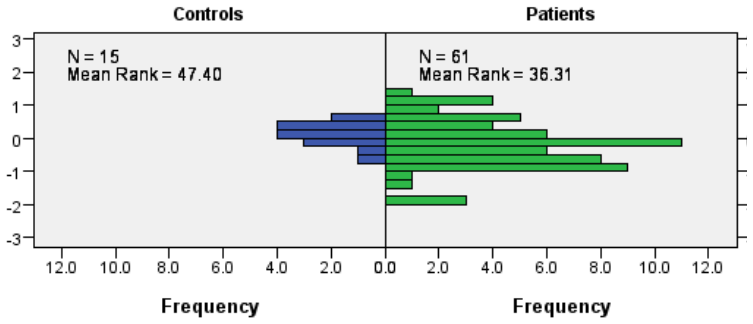
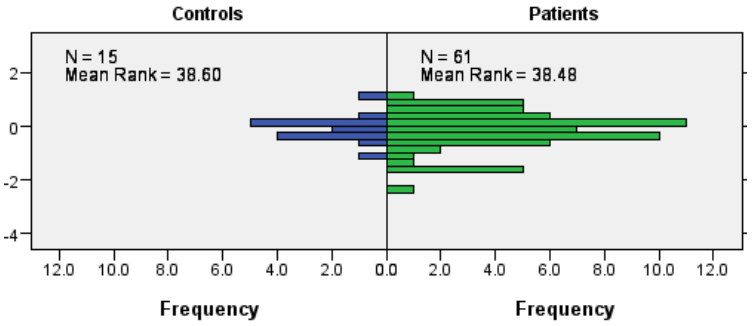
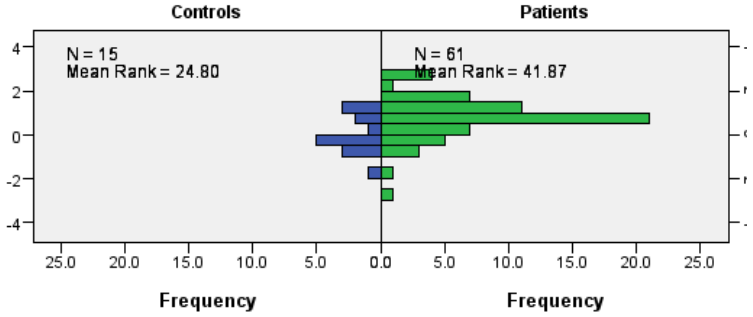
Group of tests		Difference
BRBN battery		U = 324.00, p = 0.081
Visuo-perceptual		U = 456.00, p = 0.984
Language		**U = 663.00, p = 0.007**

Figure 7.1. shows that the controls and patients have showed similar patterns of longitudinal change in their performance on the BRBN and visuospatial tests. On the language tests the controls have deteriorated more than the patients

Note. Group comparisons were performed using the Mann-Whitney U-test

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

II. Comparison of change in performance scores between the different groups of tests

In this section I've investigated the differences in how much the participants have actually changed between the three different groups of tests. First I have separately looked into the differences between the changes in performance on the three groups of tests for the control participants, to understand whether there have been any changes in performance that could potentially be attributable to other factors not related to MS, such as healthy cognitive ageing, practice effect, normal fluctuation in performance, or systematic differences in administration between researchers at phase I and phase II.

As it can be seen from the Table 7.1, the controls had shown similar patterns of change on the three groups of cognitive tests. This indicated that the differences that were observed between the patient and control participants on the change in language tests (Figure 7.1) were because the patients had improved on them, and the controls have in fact performed the same.

However, there have been differences between how much the patients have changed on the three groups of cognitive tests. The patients have on average deteriorated on the BRBN and visuoperceptual tests, and improved on the language tests, and this difference was statistically significant.

Table 7.1. Comparison of the standardized scores of change for patient and control participants on the three groups of cognitive tests

Cognitive tests	Controls (n=15)	Patients (n=61)
BRBN battery	0.09 (0.32)	-0.20 (0.73)
Visuoperceptual	-0.10 (0.49)	-0.20 (0.75)
Language	0.03 (0.80)	0.78 (0.94)
Difference	F(2) = 1.600, p = 0.449	**F(2) = 43.672, p < 0.001**

Table 7.1. shows that on average the controls haven't changed in their performance on all three groups of cognitive tests, while the patients have on average deteriorated on the BRBN and visuoperceptual tests, and improved on the language tests

Note. The differences between the groups of cognitive tests were compared using the Related-Samples Friedman's analysis of variance by ranks

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

A visual representation of the differences in standardized scores of changes between performance at phase I and performance at phase II on the three groups of cognitive tests will be discussed later on in this chapter, and can be seen in Figure 7.3.

III. Actual change in performance. Summary

After analysing the longitudinal changes in cognitive performance it was found that the controls have shown non-different amount of change on all three groups of tests. The patients have deteriorated on the on the BRBN and visuoperceptual, but not on the language tests. This indicated that the differences that were observed between the patient and control participants on the change in language tests were because the patients had improved on them, and the controls have performed the same at both phases.

7.3.2. Perceived change in performance

I. *Comparison of perceived change in performance between the patients and controls*

The next step was to investigate the patterns of how the patients and the controls have thought that they have changed. The comparison of estimations of patients' and controls' longitudinal change can be seen in Figure 7.2.

I found that the patients and the controls estimated their changes in a similar manner on the BRBN and language tests. On the BRBN tests both patients and controls more often thought that they have deteriorated, and on the language tests both patients and controls mainly thought that they haven't changed. However, on the visuoperceptual tests the patients perceived that they have deteriorated, while the controls perceived that their performance hadn't changed.

The average score of patients estimations of their change in the visuopeceptual domain was lower (Mann-Whitney $U = 624.5$, $p = 0.025$), and the difference between the distributions of patient and control estimation scores on the visuoperceptual tests was close to being statistically significant (Kolmogorov-Smirnov $Z = 1.297$, $p = 0.069$).

Figure 7.2. Comparison of self-perception of change between control and patient participants

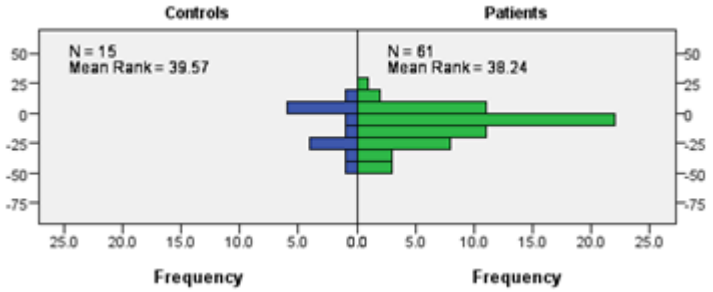
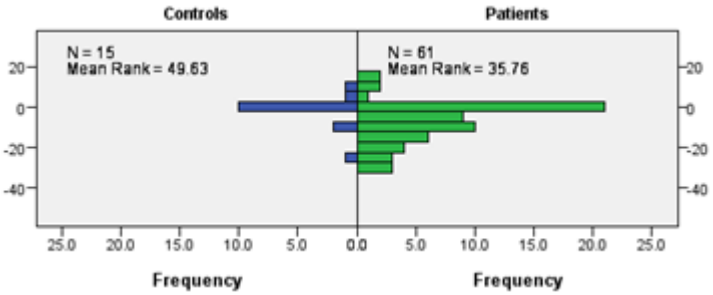
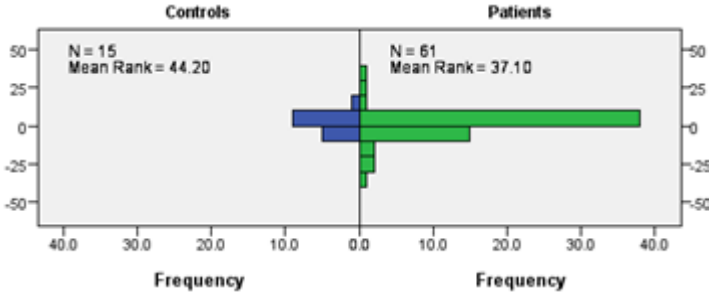
Cognitive tests		Difference
BRBN battery		U = 473.5, p = 0.833
Visuoperceptual		**U = 624.5, p = 0.025**
Language		U = 543, p = 0.223

Figure 7.2. shows that the patients and the controls have perceived their changes similarly on the BRBN and language tests, but on the visuoperceptual tests the patients thought that they have deteriorated more than the controls

Note. Positive values indicate perceived improvement, negative values indicate perceived deterioration, and values around 0 indicate perceived stability of cognitive function. The differences between the control and patient estimations were compared using the Mann-Whitney U-test

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

II. Comparison of perceived change in performance on the different groups of tests

On average, the patients have thought that they have deteriorated in their performance on all groups of cognitive tests; however, on some groups of tests they perceived to have deteriorated more than on others ($F(2) = 23.47$, $p < 0.001$). Most deterioration was perceived on the BRBN items (Figure 7.3). This tendency was also observed in the control sample, but the differences did not reach significance due to small control sample size.

Table 7.2. Comparison of self-perception of longitudinal change on the three groups of cognitive tests for patient and control participants

Cognitive tests	Controls (n=15)	Patients (n=61)
BRBN	-13.73 (18.57)	-13.16 (14.19)
Visuoperceptual	-2.07 (8.41)	-6.98 (10.44)
Language	-0.40 (8.41)	-2.74 (9.82)
Difference	$F(2) = 3.24$, $p = 0.197$	**$F(2) = 23.47$, $p < 0.001$**

Table 7.2. indicates that both controls and patients have thought that they have deteriorated on the BRBN items more than on the visuoperceptual and language items.

Note. Positive values indicate perceived improvement, negative values indicate perceived deterioration, and values around 0 indicate perceived stability of cognitive function. The differences between the groups of cognitive tests were compared using the Related-Samples Friedman's analysis of variance by ranks.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

III. Perceived change. Summary

Both controls and patients have thought that they have deteriorated on the BRBN items more than on the visuoperceptual and language items. The patients and the controls estimated their changes similarly on the BRBN and language tests, but on the visuoperceptual tests the patients thought that they have deteriorated more than the controls.

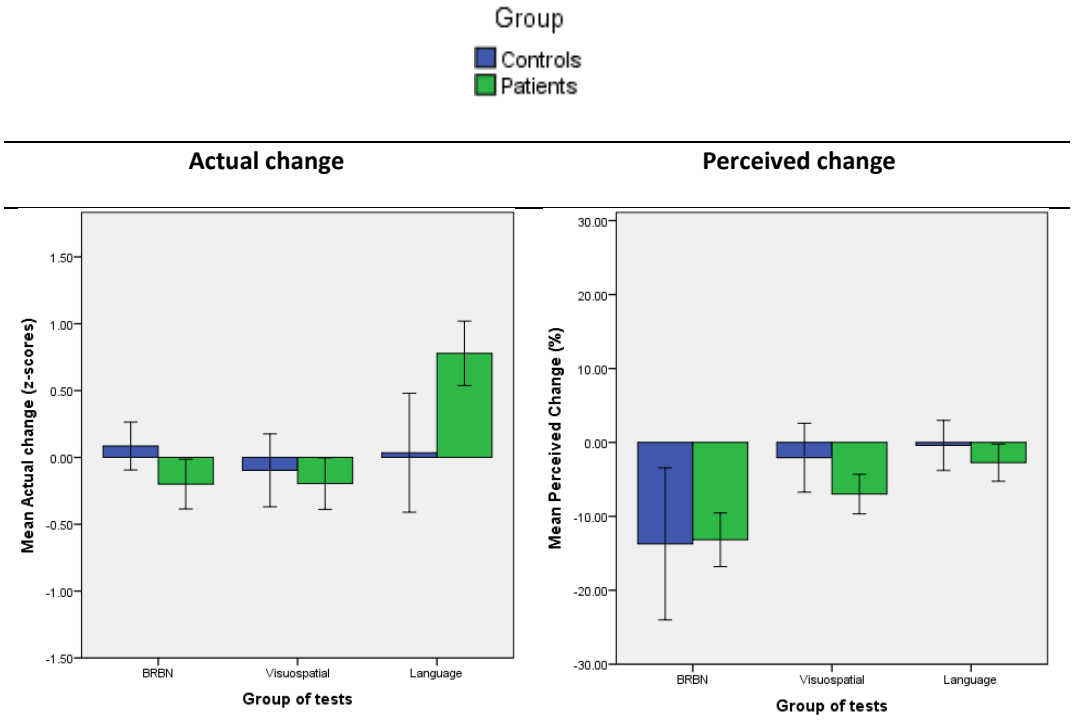
7.3.3. Relationships between actual and perceived change in performance

The next step was to investigate the trends of how the participants thought that they changed are in comparison to the trends of how they have actually changed.

I. Comparison of trends exhibited by patients and controls

The comparison between the actual and the perceived changes in performance was done by visual examination of the patterns of perceived and actual change in the control and patient samples (Figure 7.3.).

Figure 7.3. A visual representation of the differences between actual and perceived changes in performance on the three groups of tests for control (n=15) and patient (n=61) participants



Error Bars: 95% CI

Figure 7.3. shows that both patients and controls tend to underestimate their performance on all tests. A general tendency of participants to think that they have deteriorated was observed

Note. The actual change is presented in standardized scores. Positive values of perceived change indicate perceived improvement, negative values indicate perceived deterioration, and values around 0 indicate perceived stability of cognitive function.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

As it can be seen from Figure 7.3., both patients and controls tend to underestimate their performance on all tests. In other words, a general tendency to think that they have deteriorated was observed. As *on average* all participants have thought that they have deteriorated.

II. Correlations between the actual and the perceived longitudinal change

The next step was to investigate the relationship between the actual and the perceived changes in performance on the individual level, accounting for individual trajectories of change.

The first analysis included running multiple correlations between the measures of actual longitudinal change and the perceived change. As it can be seen from Table 7.3., the patients tend to be more accurate than the controls in estimating the trajectory of their cognitive change.

Table 7.3. Correlations between actual change and perceived longitudinal change

Cognitive tests	**Controls (n=15)**	**Patients (n=61)**
BRBN battery	r = 0.049, p = 0.803	**r = 0.313, p = 0.001**
Visuoperceptual	r = - 0.095, p = 0.652	**r = 0.259, p = 0.005**
Language	r = 0.041, p = 0.839	r = 0.097, p = 0.323

Table 7.3. shows that the patients have been more accurate in perceiving their cognitive changes on the BRBN battery and visuoperceptual tests

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

Figure 7.4. Relationships between actual change and perceived change on the three groups of cognitive tests

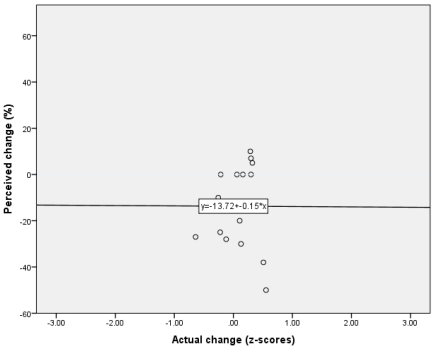
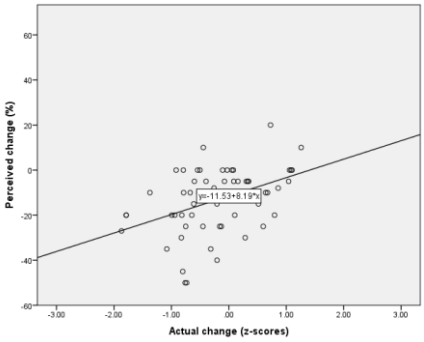
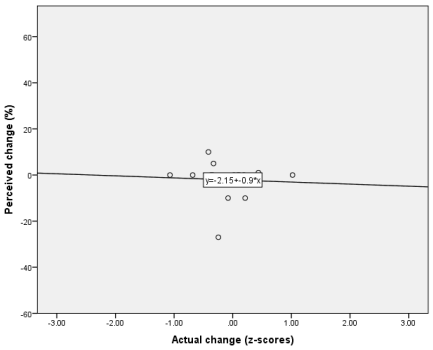
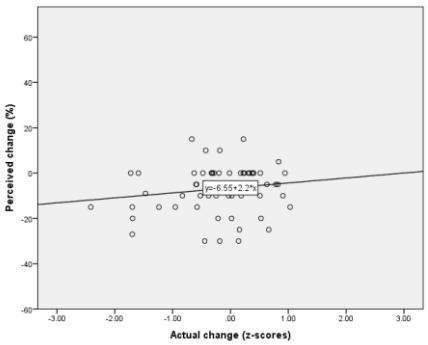
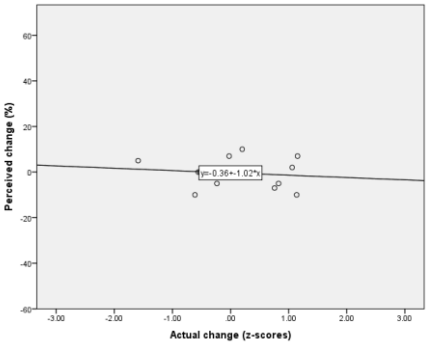
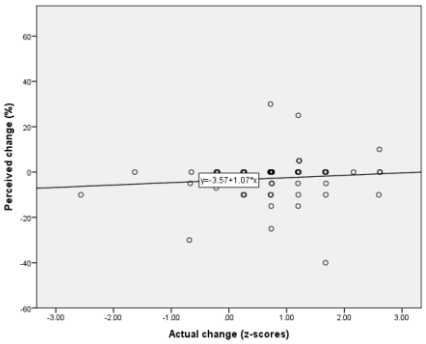
Tests	Controls (n=15)	Patients (n=61)
BRBN battery		
Visuoper-ceptual		
Langu-age		

Figure 7.4. shows that there was no relationship between the actual and the perceived change in control performance on all groups of tests, and on patient performance on the language tests. However, the patients showed some degree of accuracy in perceiving their change on the BRBN battery and visuoperceptual tests.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

III. Comparison between the actual and the perceived change in performance. Summary

A general tendency for both patient and control participants to underestimate their performance was observed on all cognitive tests. However, when compared to controls, the patients have been more accurate in perceiving their cognitive changes on the BRBN battery and visuoperceptual tests. There was no relationship between the actual and perceived change in control performance on all tests, and on patient performance on the language tests. Moreover, the patients were considered to show some level of accuracy in perceiving their change on the BRBN battery and visuoperceptual tests. This could be due to the fact that all participants think that they have deteriorated, when in reality only the patient participants have.

7.3.4. Predictors of perceived change

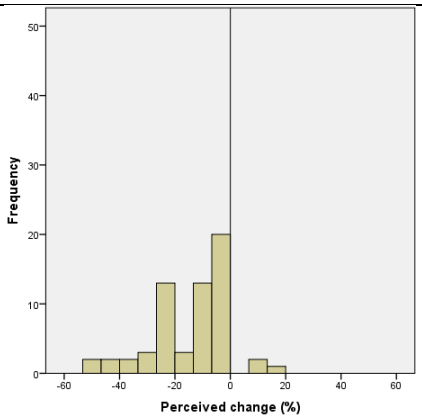
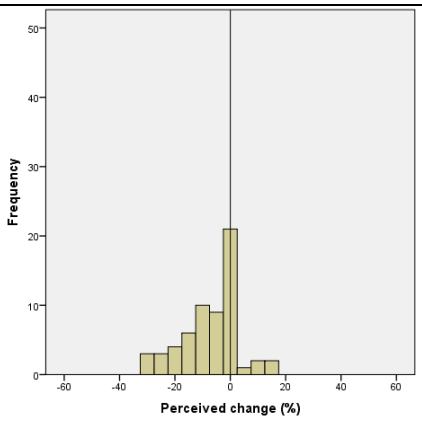
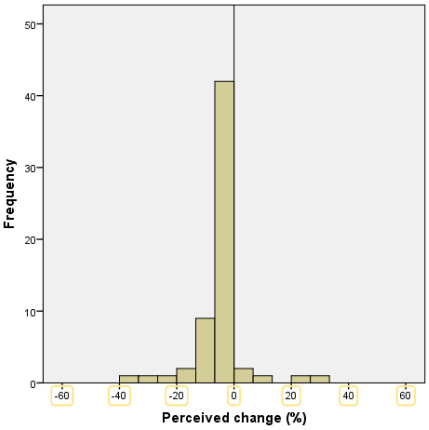
As it can be seen from previous analyses earlier in this chapter, the patient and control participants tend to perceive that they have deteriorated, regardless of what the actual change in performance was. It appeared that it was of no importance if the participants have actually improved, remained the same or deteriorated, the majority of them have thought that they have deteriorated.

The next step was to investigate what makes the patient participants think that they have changed. This was examined by running multiple linear regression analyses with the perceived change as the criterion variable and measurements of executive functioning, depressive symptomology, neurological disability, MS impact, and number of relapses in-between assessments as predictor variables.

I. Analysis of univariate normality of criterion variables

The first step before running multiple linear regression analyses was to check for univariate normality of the criterion variables. As it can be seen in Figure 7.5., the Shapiro-Wilk statistics indicated that all of the criterion variables haven't been normally distributed. However, upon visual inspection it was decided that the distributions were sufficiently normal, as all of them were unimodal with no extreme skew. Therefore it was decided to proceed with running the linear regression models.

Figure 7.5. Distributions of criterion variables for linear regression models of perceived change for patient participants

Cognitive tests	Mean (SD)	Shapiro-Wilk test	Distribution
BRBN battery	-13.16 (14.19)	W = 0.947, p = 0.011	
Visuoperceptual	-6.98 (10.44)	W = 0.931, p = 0.002	
Language	-2.74 (9.82)	W = 0.752, p < 0.001	

Note. All criterion variables were in percentages of change. Positive values of perceived change indicate perceived improvement, negative values indicate perceived deterioration, and values around 0 indicate perceived stability of cognitive function. The vertical line represents no perceived change.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological test

II. Analyses of the relationships between the criterion and predictor variables

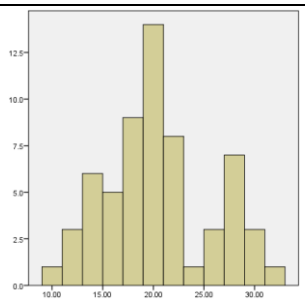
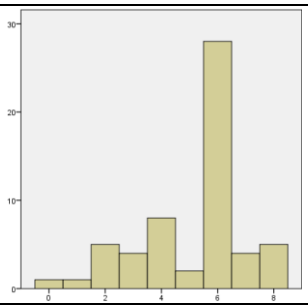
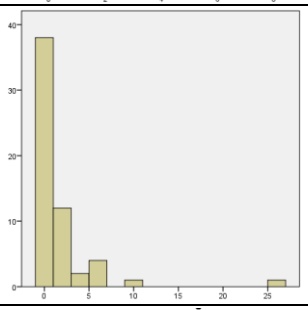
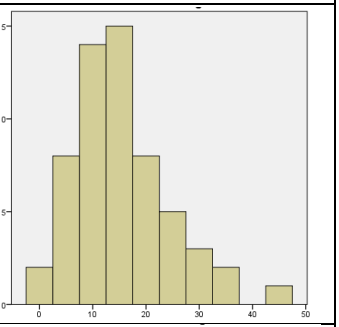
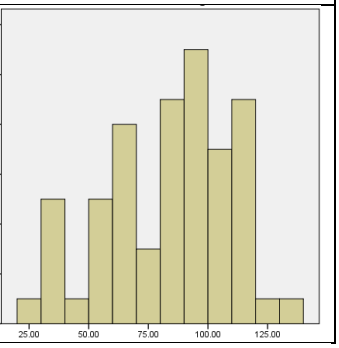
Before running the linear regression models it was necessary to identify the predictor variables worthy of including, as due to small sample sizes (61 patients) and insufficient power I could not include more than two predictor variables for the patient models.

This was done by examining the relationships between the criterion and predictor variables, and only those predictor variables that were shown to be related to the criterion variables were included in the multiple linear regression models.

The distributions of the predictor variables can be seen in Figure 7.6. It should be noted that there have been high numbers of patients with EDSS scores of 6 and zero relapses. Therefore the results of those predictors should be interpreted with caution.

As it can be seen from Table 7.4, the perceived changes on BRBN tests were associated with executive functioning, neurological disability and MS impact. The perceived changes on the visuospatial tests were associated with executive functioning. No relationships were found with the perceived change on language tests, but this was potentially caused by lack of actual change, as the patients had not exhibited notable change in their performance on the language tests.

Figure 7.6. Distributions of clinical predictor variables used in linear regression models to determine their effects on perceived change in performance for patient participants

Variable	Mean (SD)	Shapiro-Wilk test	Distribution
WLGT	20.02 (5.20)	W = 0.960, p = 0.045	
EDSS	5.19 (1.86)	W = 0.945, p = 0.01	
Number of relapses (only for RRMS and SPMS patients (n=45))	1.33 (3.68)	W = 0.877, p < 0.001	
BDI - II	15.12 (9.38)	W = 0.394, p < 0.001	
MSIS-29	82.97(25.99)	W = 0.962, p = 0.056	

Abbreviations: WLGT – Category animal fluency test, EDSS – Expanded Disability Status Scale, BDI-II – Beck's Depression Inventory 2nd Ed., MSIS-29 – Multiple Sclerosis Impact Scale.

Table 7.4. Correlations between predictor and criterion variables in models of perceived change in performance on groups of cognitive tests for patient participants

Predictor variables	BRBN battery	Visuoperceptual	Language
WLGT	**r = 0.298, p = 0.002**	**r = 0.199, p = 0.039**	r = 0.060, p = 0.550
EDSS	**r = - 0.247, p = 0.011**	r = -0.054, p = 0.586	r = -0.169, p = 0.101
Number of relapses	r = 0.061, p = 0.561	r = -0.022, p = 0.836	r = -0.070, p = 0.528
BDI-II	r = -0.062, p = 0.508	r = -0.119, p = 0.212	r = 0.042, p = 0.673
MSIS-29	**r = -0.200, p = 0.030**	r = -0.088, p = 0.351	r = -0.132, p = 0.180

Table 7.4. shows that the perceived changes on BRBN tests were associated with executive functioning, neurological disability and MS impact. The perceived changes on the visuoperceptual tests were associated with executive functioning

Note. The relationships were examined employing the Kendall Tau correlation

Abbreviations: WLGT – Category animal fluency test, EDSS – Expanded Disability Status Scale, BDI-II – Beck's Depression Inventory 2nd Ed., MSIS-29 – Multiple Sclerosis Impact scale

Based on the findings presented in Table 7.4. it was proceeded with running a multiple linear regression model with three predictor variables for the perceived change on the BRBN tests, and a simple linear regression model with sole predictor for the perceived change on the visuoperceptual tests. It was decided not to continue with further analyses to explain perceived changes on the language tests.

I. Linear regression analyses to investigate the predictors of perceived change in performance on the cognitive tests

a) BRBN tests

Based on the presented associations (Table 7.4.), the following linear regression model was proposed in order to identify the predictors of perceived change in performance on the BRBN tests.

$$\textit{Self-awareness of longitudinal change on BRBN tests} = \alpha + \beta_1 (\textit{executive functioning}) + \beta_2 (\textit{neurological disability}) + \beta_3 (\textit{MS impact})$$

Before running this model the predictor variables were checked for multicollinearity. WLGT did not correlate with EDSS (r = -0.113, p = 0.231, Kendall Tau, 2-tailed) and MSIS-29 (r = -0.152, p = 0.093, Kendall Tau, 2-tailed). However, EDSS and MSIS-29 scores were moderately correlated (r = 0.444, p < 0.001, Kendall Tau, 2-tailed). This auto-correlation was expected, as neurological disability is the major factor contributing to MS impact.

However, even though much of EDSS and MSIS-29 variance was shown to overlap, these two factors could not be considered equal, as physical disability is not the sole predictor of MS impact. Emotional, sleep, and quality of life factors are also components of the MSIS-29, and I believe that they play a significant role as well in perceiving changes in cognitive functioning.

Instead of excluding one of these variables out of the equation it was decided to run two equations instead, in one of them including WLGT and EDSS, and in another including WLGT and MSIS-29, with the aim of identifying the model with a better fit.

The following two models are presented below:

$$\textit{Self-awareness of longitudinal change on BRBN tests} = \alpha + \beta_1 \textit{(executive functioning)} + \beta_2 \textit{(neurological disability)}$$

and

$$\textit{Self-awareness of longitudinal change on BRBN tests} = \alpha + \beta_1 \textit{(executive functioning)} + \beta_2 \textit{(MS impact)}$$

For the first model with WLGT and EDSS as predictor variables the Durbin-Watson test values indicated that there was no autocorrelation (d = 1.765), and it was concluded that it's safe to proceed with the multiple linear regression model.

When perceived change on the BRBN tests was predicted it was found that both WLGT score (Beta = 0.325, p = 0.006) and EDSS score (Beta = - 0.362, p = 0.002) were significant predictors. The overall model fit was $R^2_{adj.} = 0.256$, $p < 0.001$ (Figure 7.7).

Figure 7.7. The effects of executive functioning, neurological disability, and MS impact on the perceived change in performance on the BRBN tests for patient participants

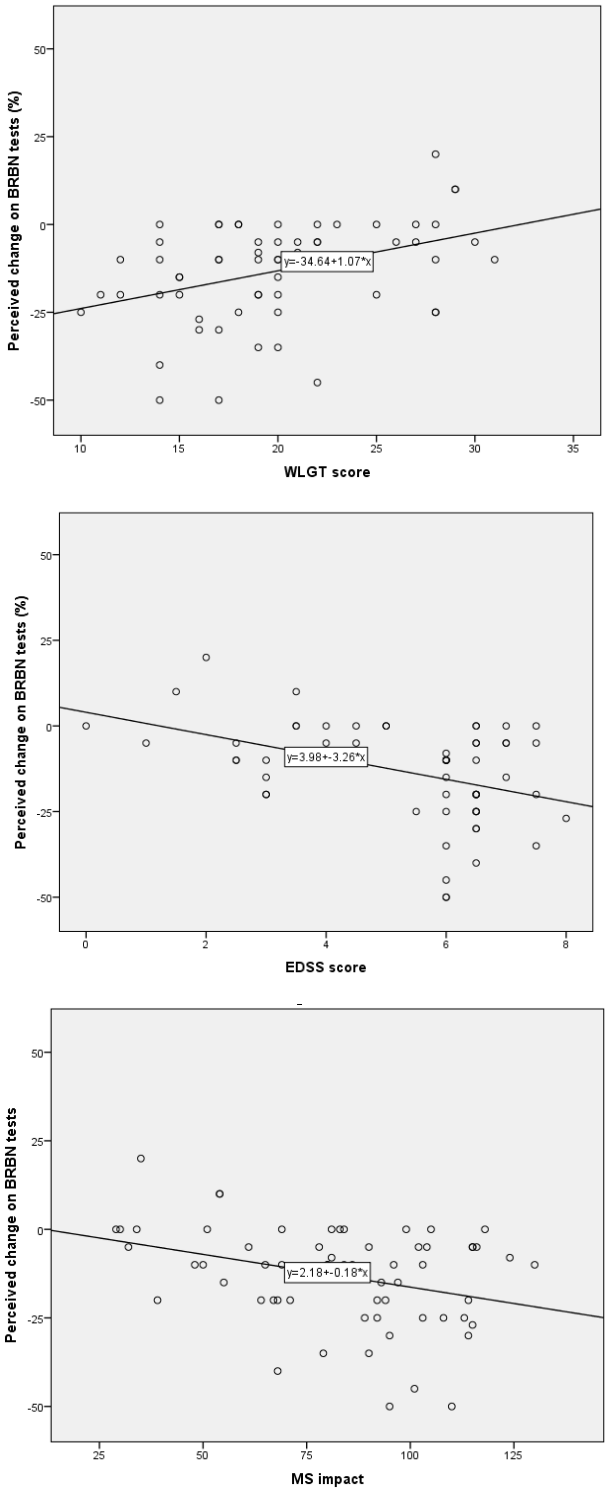


Figure 7.7. shows that executive dysfunction, neurological disability, and MS impact were predictive of perceived deterioration in performance on BRBN tests

Abbreviations: BRBN – Brief-Repeatable battery of neuropsychological tests, WLGT – Category animal fluency task, EDSS – Expanded Disability Status Scale

For the second model with WLGT and MSIS-29 as predictor variables the Durbin-Watson test values indicated that there was no autocorrelation (d = 1.829), and it was concluded that it's safe to proceed with the multiple linear regression model.

When perceived change on the BRBN tests was predicted it was found that both WLGT score (Beta = 0.338, p = 0.006) and EDSS score (Beta = - 0.270, p = 0.026) were significant predictors. The overall model fit was $R^2_{adj.} = 0.198$, p = 0.001.

From running both of these models it can be concluded that executive dysfunction, neurological disability, and MS impact are predictive of perceived deterioration in performance on BRBN tests, although neurological disability could be considered to be a better predictor of perceived decline than MS impact.

b) Visuoperceptual tests

Based on the presented associations (Table 7.4.), the following linear regression model was proposed in order to identify the predictors of perceived change in performance on the visuoperceptual tests.

Self-awareness of longitudinal change on visuoperceptual tests $= \alpha + \beta_1$ *(executive functioning)*

When perceived change on the visuoperceptual tests was predicted it was found that the WLGT score (Beta = 0.251, p = 0.030) was a significant predictor. The overall model fit was $R^2_{adj.} = 0.062$, p = 0.030. A visual representation of the relationship between WLGT and perceived change on visuoperceptual tests can be seen in Figure 7.8.

Figure 7.8. The effects of executive functioning on the perceived change in performance on the visuoperceptual tests for patient participants

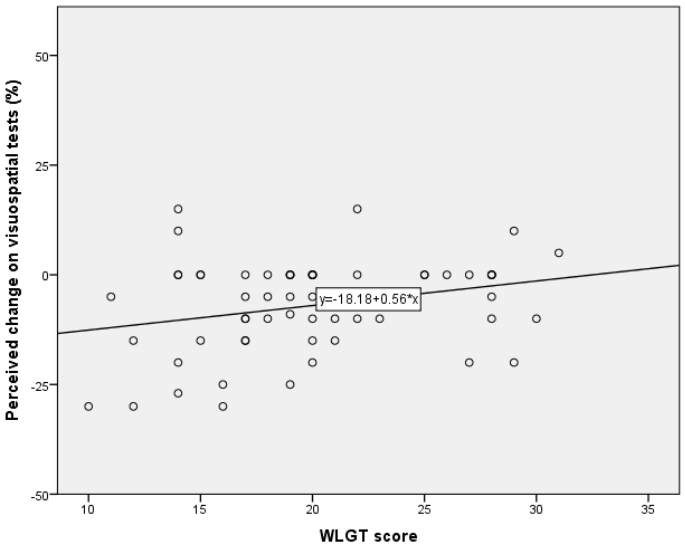


Figure 7.8. shows that executive dysfunction was predictive of perceived deterioration in performance on the visuoperceptual tests

Abbreviations: WLGT – Category animal fluency task

7.3.5. Predictors of perceived change. Summary

From running linear regression analyses I've found that executive dysfunction, neurological disability, and MS impact were predictive of perceived deterioration in performance on BRBN tests, and that executive dysfunction was the sole predictor of perceived deterioration in performance on the visuoperceptual tests.

7.4. Chapter summary

After analysing the longitudinal changes in cognitive performance it was found that the controls have equally deteriorated on all three groups of tests. The patients have deteriorated on the on the BRBN and visuoperceptual, but not on the language tests. This indicated that the differences that were observed between the patient and control participants on the change in language tests were because the patients had improved on them, and the controls have performed the same at both phases.

Both controls and patients have thought that they have deteriorated on the BRBN items more than on the visuoperceptual and language items. The patients and the controls estimated their changes similarly on the BRBN and language tests, but on the visuoperceptual tests the patients thought that they have deteriorated more than the controls.

A general tendency for both patients and controls to underestimate their changes in performance was observed on all cognitive tests. However, when compared to controls the patients have been more accurate in perceiving their cognitive changes on the BRBN battery and visuoperceptual tests. This could be due to the fact that all participants think that they have deteriorated, when in reality only the patient participants have. There was no relationship between the actual and perceived change in control performance on all tests, and on patient performance on the language tests. However, the patients were accurate in perceiving their change on the BRBN battery and visuoperceptual tests.

From running linear regression analyses I've found that executive dysfunction, neurological disability, and MS impact were predictive of perceived deterioration in performance on BRBN tests, and that executive dysfunction was the sole predictor of perceived deterioration in performance on the visuospatial tests.

The analyses performed in this chapter have fulfilled the chapter aim of investigating the patients' self-estimates of the trajectory and extent of their cognitive change, and identifying the factors that have an effect on those self-estimations.

Chapter Eight. Discussion

8.1. Purpose of the study and chapter overview

The aim of this study was to investigate the longitudinal changes in cognition in MS, by analysing the pattern of how MS-related cognitive decline starts and spreads, and the role that the clinical and non-clinical variables play on developing new cognitive deficits. In addition, this study was the first to systematically examine the patients' perception of cognitive decline in MS samples, and related their self-perception of change to the actual longitudinal changes in performance on neuropsychological tests. This project has successfully addressed the aims and answered the research questions raised in the introduction.

In this chapter the results from the longitudinal study were explained. The main body of the discussion chapter was composed of two parts, the first part corresponding to the interpretation of the study findings, and the second part was allocated for critical evaluation of the methodology, clinical application of the results, and guidelines for future research.

8.2. Interpretation of the findings

8.2.1. Dimensionality of cognitive impairment

The first aim of this project was to investigate the dimensionality of MS-related cognitive impairment. It was hypothesized that MS affects all cognitive domains, although independently and not necessarily at the same frequency. Another hypothesis for longitudinal changes was made as well, where it was predicted that the pattern of cognitive deficits would become more homogeneous over time at the population level, reflecting increasing multidimensionality at the individual level.

This study has successfully addressed the two aims. Based on the analyses of prevalence of cognitive deficits in Chapter Five it was shown that indeed as predicted the cognitive deficits can be observed in all cognitive domains and that no functions were spared. With respect to the longitudinal examination, the cognitive deficits initially appear heterogeneous on individual basis, although with

MS progression, and development of new impairments (mainly in the areas of information processing speed and visuoperceptual functions) the individual patterns of deficits had the tendency to become more homogenous. This indicated a reduction in the dimensionality of the MS population with respect to their pattern of cognitive impairments over time. At phase I the patients classified into multiple groups based on their performance, indicative of the many ways in which cognitive deficits start (apparently in random and typically isolated cognitive impairments). However, three years later (at phase II), only two clusters were present; this showed that with acquirement of new deficits the patient cohort had become more homogeneous in their expression of their cognitive symptoms. These results remained unchanged after addressing missing observations using the multiple imputation technique (Appendix E).

These findings would imply that MS affects individuals by initially causing a small set of domain-specific impairments (that differ between patients), and later evolves into (a larger) multi-domain set of impairments at the individual level. Therefore it can be concluded that both hypotheses about the dimensionality of MS related cognitive impairment were supported in this study.

These findings tend to contradict those reported by previous studies that argued that in pwMS cognitive deficits start in certain cognitive functions and then spread to other cognitive functions (such as Denney et al., 2008; Morrow et al., 2009; Glanz et al., 2012). From the results of this study it appears that the cognitive deficits that the pwMS pick up are sporadic in their nature, and could not be narrowed down into particular patterns of how they emerge. However, my longitudinal analyses suggested that the cognitive deficits in pwMS slowly accumulated, spreading into processing speed and visuoperceptual domains, leading to development of major problems with longer disease duration.

However, besides the duration of MS, it seemed that there was no other reason for new deficit acquirement and the population becoming more homogeneous. Besides some individuals with SPMS who were in general much worse off than the rest of the sample (e.g. individuals with EDSS score of 8 and above), I could not identify a group of individuals who suddenly flipped into progression and exhibited severe cognitive decline. Again this contradicts previous literature that claim that specific groups of pwMS can be identified (such as ‘cognitively impaired’ and ‘cognitively

unimpaired') at baseline, and that those groupings could be predictive of future cognitive decline (Schwid et al., 2007; Deloire et al., 2010; Kujala et al., 1997). This was not observed in this sample.

This study can be deemed suitable for assessing the dimensionality of cognitive deficits in pwMS due to several reasons. Firstly, in my battery I've included measurements of a broad range of cognitive functions, while in previous research the test batteries employed were heavily weighted towards memory, attention and processing speed functions. Secondly, I had investigated all three MS types in my sample, and their recruitment rates were reflective of the overall PPMS, SPMS and RRMS prevalence in MS population. This implies that the results of this study can be generalizable to the overall MS population, when the results from most previous studies can only be generalizable to certain subsamples of pwMS. Moreover, in this project I managed to secure high follow-up rates, therefore the results of longitudinal changes in domain-specificity can be deemed representative as well.

After having analysed the longitudinal changes in the dimensionality of cognitive deficits in pwMS the next step was to try to explain what causes them.

8.2.2. Predictors of cognitive change

The second aim of the study was to examine the predictors of longitudinal change in cognition in pwMS. It was hypothesized that age, gender, pre-morbid IQ (involving pre-morbid cognitive leisure and educational exposure), affective comorbidity, and drugs, would all have an effect on the longitudinal changes in cognitive functioning. This hypothesis was tested by running separate linear regression models (Chapter Six) to investigate whether these factors have an effect on cognitive changes in five groups of cognitive tests: verbal memory, visuospatial memory, processing speed, visuoperceptual and language. The research aims could be considered to have been addressed successfully and interestingly it was found that the factors that influenced cognitive decline were specific for each of the cognitive domains.

The change in verbal memory domain was linked to the levels of neurological disability and depressiveness. The change in visuospatial memory didn't seem to be linked to any clinical variables. The change in processing speed was linked to MS type only (SPMS, and to a lesser extent PPMS).

The change in visuoperceptual domain was linked to neurological disability, MS type and MS modifying medication. The change in the language domain was linked to the level of depressiveness. It appears that the heterogeneity of causes of MS-related cognitive decline reported by previous studies could be explained by which cognitive tests were used during their assessments, as decline in performance on different tests was explained by different clinical variables in this study. However, summing up, neurological disability, MS type and levels of depressiveness were the most common predictors of change in cognitive functioning. These findings have been thoroughly described elsewhere (Patti et al., 1998; Amett, 2005; Amato et al., 2010).

The MS modifying medication had an effect only on one group of cognitive tests, and this could be because the way in how the drugs affect cognition is still unclear. Depending on the individual and on the actual drug, the disease modifying drugs may have no effect, a deleterious effect, and an advantageous effect on cognition. This could potentially explain the lack of clarity of the MS modifying treatment effect on cognitive changes in this sample.

In contrast to the effect of clinical variables, the demographic and cognitive reserve variables did not seem to have predictive value on the cognitive changes. Once again this could be used to question the arguments found in the literature that some individuals with MS are more robust to cognitive decline than others (Deloire et al., 2010; Camp et al., 2005; Kujala et al., 1997).

I. Potential explanations of why some of expected effect was not found

Even though based on the previous literature this model was expected to explain the longitudinal changes in cognition, several potential explanations for why some of the predictor factors did not have an effect on our dataset were identified.

In the model it was expected that in pwMS age adds pathology (plus interacts with comorbidity) and the effects of age would tend to intertwine with the effects of disease duration. Even though it was expected for older age to have an effect on the brain function, in our sample there were also individuals with late onset MS (older age fewer years with MS) who had relatively good cognitive outcome, and individuals with early onset MS (younger age but more years with MS) who had worse cognitive outcome. This could potentially explain why no effect of age was observed in our sample.

However, the findings may be different if everyone in our sample was followed up from the point of diagnosis, as the participants in this study have had very variable age and duration of MS.

Even though previous literature has reported that men are more likely to experience cognitive decline, in this study no specific effect of gender was found. There is some literature on gender-based differences in specific cognitive functions, such as men performing better at tasks of information processing speed and visuospatial functions, and women performing better on memory tasks (Halpern, 2013); and it could be that the gender-specific effects of MS could have been masked by these differences in abilities.

The lack of effect of premorbid functioning in this dataset did not support the hypotheses, and this could have been caused by several reasons. To begin with, as has been mentioned in the introduction, cognitive reserve related compensation is reported to be most beneficial at earliest stages of MS and may, however, fail with progression of damage (Amato et al., 2013). As this study included participants with variable durations of MS, it may be that there could have been more effect of premorbid reserve on some patients than the others. However, this question was not analysed in more detail as the primary goal was to keep consistency in the methodology of predictor analysis, and therefore to run the same analyses for all of the predictor variables. This way it was allowed for comparisons between the effects of different predictors to be made.

The second reason why no effect of premorbid intelligence was observed could be because the measurements of pre-morbid IQ employed in clinical research could be considered to be able to give only a limited indication of the actual pre-morbid performance. The tests of premorbid functioning are tests of semantic storage and were created for neuropsychological assessments of groups of patients with more homogeneous locations of brain pathology (such as AD or FTD), whereas due to heterogeneity of brain lesion locations in MS samples, some pwMS may be more in advantage than others when performing on the tests of premorbid intelligence. This could in part explain the lack of effect of premorbid abilities on longitudinal change in cognitive functioning observed in this sample.

In this study the protective effect of physical exercise may have been limited, as the means of collecting such information in this study may have been overly simplistic. The methodology of

collecting information about exercising employed in this study was coded binary as '0' 'no exercise' and '1' 'exercising'. Since many types of physical exercise exist (from easy to strenuous) an argument can be raised about the level of their comparability. It could be that choosing to code the levels of exercise on a 5-point or 7-point Likert scale rather than binary could be considered to have been a more suitable approach for this study. Moreover, the other reason why no link between exercising and cognitive change was observed could be because the ability to exercise was masked by the levels of neurological disability, as in some cases the physical disability resulting from chronic MS progression (such as dizziness or tremor) may be the cause preventing some individuals from engaging in exercise more than others. Therefore it could be considered, that the relationship is actually the other way round from what was expected, and that in this instance when working with MS samples, exercising could be regarded as a derivative from the progression of the disease, rather than from personality traits such as health-consciousness or positivism.

To sum up, out of the many potential predictor variables studied, in this study the longitudinal changes in cognitive functioning seemed to be linked only to the levels of neurological disability and depressiveness, and to some extent to MS type and MS modifying treatment uptake. The other hypothesized predictor variables seemed to have no effect on the cognitive decline. The next step was to investigate whether the patients themselves were aware of their longitudinal changes in cognition, and what factors had influence on making them perceive that their cognitive functioning has deteriorated.

8.2.3. Self-perception of change

The third aim of this project was to investigate the patients' perception of their deficit progression. This aim was successfully fulfilled by comparing the patients' actual changes in performance with their perceived changes.

A general tendency for both patients and controls to underestimate their performance was observed on all groups of cognitive tests. However, when compared to controls the patients have been more accurate in perceiving their cognitive changes on the BRBN battery and visuoperceptual, but not the language tests. This could be due to the fact that all participants have thought that they have deteriorated, when in reality only the patient participants have. Moreover, since in this project the

pwMS have been shown to perceive their changes similarly to the controls, it could be argued that the tendency observed in frontal-executive disorders (such as FTD) where the patients underestimate or fail to acknowledge their deficits, was not observed in MS.

In this study the factors that had an effect of self-estimations of perceived change for BRBN tests were executive dysfunction, neurological disability, and MS impact. Executive dysfunction was the sole predictor of perceived deterioration in performance on the visuoperceptual tests, and none of the factors were predictive of perceived change on the language tests.

However, it could be considered that the observed predictive value of executive functioning could have been to some extent influenced and inflated by the choice of the test used to assess the executive functions, indicating that the WLGT may not have been the most suitable variant for this purpose. WLGT is considered to be a category naming test, and not all of the sources of previous literature accept it as a measure of executive functioning (Phillips, 1997). Nonetheless, since the original phase I battery did not include a more suitable measure of executive functions, for the purposes of this study it was chosen to employ the closest equivalent of an executive function assessment.

As hypothesized, the proposed model had the best fit for the BRBN items, and could be considered unsuitable for the visuoperceptual and language items. This could be because the model was built based on the previous literature where all the previous work on the awareness of deficits in MS has been conducted employing the BRBN battery tests. It can be considered that separate models should be built for the perception of visuoperceptual and language difficulties, and that could include measurements of visual or contrast acuity for visuoperceptual functions, and speech or vocabulary measurements for language functions. It can be assumed that those measurements would be more predictive of perceived decline on the visuoperceptual and language abilities.

I. Controls and insight

Even though the focus of this investigation was around self-evaluations in the patient cohort, however, through additional analyses we found that the controls had the tendency to underestimate their performance when evaluated cross-sectionally at phase II. This was an important observation as so far we did not find any work concentrating on the phenomenon of the under-estimations of their

cognitive performance of the general population. The only published work that we could find on the general population was in relation to estimating self-worth, rather than evaluating cognitive performance, which is an entirely different concept. Only one study has measured self-estimations of cognitive performance in healthy adults (although only used them as controls as their focus was on FTD, CBD and PSP patients) and that study has also reported that the controls had the tendency to underestimate their performance (O'Keefe et al., 2010). That study had also employed composite scores and similar statistical methods to ours to define accuracy of self-estimations. However, besides stating the finding, they did not emphasize on it or try to explain it. What they observed O'Keefe et al. called 'normal estimations' rather than 'underestimation', and they didn't look into them separately, just as a reference point for comparison in their analyses of patient estimations.

Therefore it could be concluded that the general population do not represent the perfect reference point for comparisons, as they exhibit a natural tendency to underestimate their own performance. Therefore future studies comparing self-estimations between groups of people should take that into account. Moreover, this phenomenon needs to be explained and investigated in more detail in future work, as interpretation of self-estimations in the control sample is not as straightforward concept as it had been thought. Only the understanding of the causes behind underestimations in healthy self-evaluations could lead to a better understanding of the impaired self-evaluations and insight deficits in the patient samples.

II. PwMS and insight

In this study we found that the patient participants showed a similar pattern as the controls in estimating their performance at phase II, and as a group did not exhibit clear deficits in insight common to other neurological conditions that involve damage to the frontal cortex (O'Keefe et al., 2010). In general, we found that those pwMS who had cognitive deficits were able to identify their presence and their degree on all groups of cognitive tests, and, similarly as the healthy controls, they had a tendency to underestimate their performance. This trend was more evident on BRBN and visuoperceptual tests, and less evident on language tests, where both patients and controls provided estimations of a wide range of levels of accuracy. Moreover, there was a subgroup of pwMS that overestimated their performance, and that small subgroup could be defined by poorer scores on

frontal-executive tests. Therefore it can be concluded that even though pwMS exhibited deficits on tests of frontal-executive functions, they, as a group, were not prone to awareness disorders and were overall able to evaluate their performance with the same level of accuracy as the controls. Future studies could compare pwMS to other neurological conditions that exhibit frontal-executive problems (FTD, MND, PSP, CBD, etc.) as our study has shown that even though frontal-executive dysfunction is linked to higher levels of awareness deficits, but it is not the sole predictor of disordered insight. Therefore it can be concluded that there is basis for other variables that cause awareness problems, or, perhaps, the degree of frontal-executive deficits was much milder in our MS sample than it was in previous insight studies into other disorders. Moreover, it could also be that the awareness deficits develop only once frontal-executive problems have reached a certain threshold, which could explain the lack of linear relationship between the variables.

In this study we found that pwMS were generally accurate in estimating their current cognitive status, therefore it should be safe to assume that in clinical practice, when evaluating cerebral functions as part of EDSS, the patients should be considered accurate in their reports of cognitive functioning (or at least no major discrepancies were shown in this study to support otherwise). However, since pwMS could be considered accurate in their cognitive estimations, in both research and clinical practice it could be beneficial to supplement the EDSS with Multiple Sclerosis Neuropsychological Questionnaire (MSNQ; Benedict et al., 2003). MSNQ is a short 15-item self-report screening measure of neuropsychological functioning that includes questions addressing cognitive domains often disrupted in MS. MSNQ could be a particularly useful instrument if the investigator is interested in collecting more information about multiple cognitive functions, and this could be done both at baseline and at follow-up in order to evaluate longitudinal change, either as part of an observational or an interventional study. Due to it's multidimensionality the MSNQ would yield a more detailed report of perceived cognitive functioning than that currently collected via EDSS. Moreover, the MSNQ is a well-validated measure that also gives insight into the partner's or carer's perspective of the patient's cognitive status, therefore overall could be considered to be a superior method for collecting self-estimations to that which has been used in this study.

In our study, however, we found that as a group the patients were accurate not only when estimating their cognition cross-sectionally, but also when estimating the longitudinal change. This contrasts previous findings from one study where pwMS have been found to not be fully aware of their current cognitive deficits (or not report them reliably) when examined cross-sectionally, however, when examined longitudinally, pwMS were found to be accurate in their reports of perceived levels of cognitive change (Christodoulou et al., 2005). Interestingly, in their study the authors report overestimations, and not underestimations of performance, which is surprising to some extent, as their study had employed the same cognitive tests (BRBN battery). However, the method of collecting the awareness data selected for that study was different – in their study the 5-item Perceived Deficits Questionnaire and a 2-item memory and attention/concentration questionnaires were used to assess self-perceived cognitive impairment. Estimating one's level of cognitive deficits is a highly subjective task, prone to be affected by how the question is formulated. Therefore it could be concluded that the discrepancy of findings on awareness between this study and previous research could be explained by the methodology behind awareness data collection.

From the analysis of self-estimations of performance at phase II we found that overall, similarly as the controls, the patients were prone to underestimate their performance, although there was a range in accuracy, including a number of patients who have over-estimated their performance. This has been reported by another study that investigated the self-perception of cognitive deficits in MS, which has shown that the pwMS could be classified according to whether they overestimated or underestimated their cognitive ability (Carone, Benedict, Fishman, & Weinstock-Guttman, 2005). However, instead of comparing the patients' self-reports to their actual performance, that study compared them to informant ratings (usually those of a spouse or a family member). In their study, compared to underestimations, overestimations of performance were linked to lower levels of depression and conscientiousness, and greater degrees of cognitive impairment, euphoric behavioural disinhibition, and unemployment (Carone, Benedict, Fishman, & Weinstock-Guttman, 2005). This is in line with our findings, which showed that in the very few cases that exhibited overestimations, they were associated with higher levels of MS impact and neurological disability, but not depression. This highlights the behavioural consequences of insight deficits in a subsample of MS patients, and invites to further study insight focusing on the peculiarities of this small group of pwMS.

8.2.4. MS type and cognition

The following section revolves around the generalizability and implications this work has in terms of understanding cognition in different types of MS. Our findings are compared to those of previous studies with the aim of understanding the cognitive features associated with progressive and relapsing-remitting courses of MS.

I. Differences in performance of MS subgroups on cognitive tests

Studies comparing the cognitive profiles of patients with three types of MS have been quite recent (Denney, Sworowski and Lynch, 2005). Historically the more common practice has been to compare RRMS to chronic progressive patients (combining PPMS and SPMS subtypes together). Typically, in those studies patients with chronic progressive MS are found to have a more pronounced cognitive impairment than those with relapsing-remitting disease (Beatty et al., 1988, Beatty et al., 1989, Kujala et al., 1995).

In the few studies where patients with PPMS and SPMS have been distinguished, most investigators have reported greater overall cognitive impairment in those with SPMS (D'Amico et al., 2016; Huijbregts et al., 2006), athough others argued that PPMS and SPMS patients couldn't be differentiated by their neuropsychological performance once matched for disease duration and disability (Foong et al., 2000).

In our study we found that the SPMS patients showed worse overall cognitive performance, and had failed the cognitive tests more often than the RRMS and PPMS patients on all cognitive domains. Regarding the differences between the RRMS and PPMS, in general these two groups of patients had exhibited very similar cognitive profiles. In those cases where there had been any differences between these subtypes at phase I, they disappeared when re-assessed three years later at phase II. In our study no crude differences between these subtypes were found, only minor, where PPMS showed slightly more deficits on information processing speed and visuospatial memory tasks.

Other studies report similar tendencies with the mildest cognitive impairments in RRMS, slightly more common impairments in PPMS and most prevalent and global, multi-domain cognitive

impairments in SPMS (Beatty et al., 1989; Comi et al., 1995; Filippi et al., 1994; Huijbregts et al., 2006; Ruet et al., 2013). However, it should be taken into consideration that those studies did not investigate all three groups of patients together using the same tests. Even though these findings have been shown to some extent elsewhere, in this study we were able to show them all in one cohort and with the same tests administered to everyone.

There had been attempts in previous research to explain poorer cognitive performance in SPMS patients. SPMS can be considered a more advanced stage of MS, one that might be expected to be present with greater cognitive impairment. Such patients have a prior history of relapsing disease that has subsequently devolved to a progressive course. In contrast, PPMS patients have no prior history of relapsing disease and therefore a comparison between them and relapsing patients does not necessarily entail a difference in chronicity (Denney et al., 2005). Overall RRMS patients are relatively better off, which is not surprising as they are, on the average, characterised by younger age and shorter disease duration.

However, the differences between the MS subtypes observed in this and in earlier studies were shown to be robust in a very recent study, where after controlling for age, sex, disability level, disease duration, and education level, patients with SPMS were at least twofold more frequently impaired than patients with late RRMS in information processing speed, executive functions, verbal fluency, verbal episodic memory, working memory, and visuospatial construction (Planche et al., 2016). This illustrates that the pattern of differences between MS subtypes observed in this study has been shown to be robust. Even after removing the effects of potential confounding variables the SPMS remained the most cognitively impaired subgroup of pwMS.

Little research has been done in order to investigate the differences among the MS subtypes in each cognitive domain. In the following sections our findings from each cognitive domain on group differences among patients with different subtypes of MS will be discussed.

II. Language

In our study we found language impairments to be rare in pwMS. At phase I only three and at phase II none of the 82 pwMS studied had exhibited problems in the language domain. The language impairments remained rare even after accounting for the missing observations.

The three patients that had failed both of the language tests at phase I were all SPMS type. However, when studied separately, the SPMS patients had a higher tendency to fail the test of reception of grammar (TROG), but not the test of spontaneous language production (BCT). Of those who failed the TROG, 7 were SPMS (19% of SPMS sample), 2 were PPMS (8% of PPMS patients), and 3 were RRMS (6% of RRMS sample). Of those who had failed the BCT, 3 were SPMS (8% of SPMS sample), 5 were PPMS (19% of PPMS sample), and 9 were RRMS (18% of RRMS sample). Even though it has been shown with this study that isolated deficits on a task measuring spontaneous language production appear more common in PPMS and RRMS patients, however, they do not tend to form a more global language deficit, as they do in SPMS. Moreover, for the RRMS and PPMS patients the deficits on the BCT to some extent might have been an artefact caused by the flawed methodology of BCT administration and scoring (as explained in more detail later on in the discussion).

Language cognition has not been extensively studied in MS samples. Most of previous research revolves around information processing speed, memory and attention functions; however, the few studies that have investigated language deficits in MS have shown similar results to ours. In their study Mackenzie & Green (2009) have also shown that chronic progressive MS patients exhibit patterns of global deterioration in the language domain, encompassing both expression and comprehension. In their study Mackenzie & Green have studied the chronic progressive MS patients as a group, thus they did not compare the performance of PPMS with those with SPMS. Therefore our study has added to their findings that more global deterioration in language, as a result of global cognitive decline, is present in SPMS patients and not in PPMS patients.

The assessment battery employed in Mackenzie & Green's study contained tests that measured only the language functions. In our study we have studied language in relation to other cognitive functions.

We found that language deficits do exist, but not as isolated deficits, as they were shown to emerge only in combination with other cognitive difficulties. It could be considered that impairments in language are a result not only of dysarthria (Mackenzie & Green, 2009), but also of a more global cognitive decline, both associated with SPMS course. It could be speculated that the language impairments in MS patients tend to predominantly be caused by initial learning, memory and information processing problems, rather than by isolated language deficits.

Our findings contradict those observed by Connick et al. (2013) who, by studying people with progressive MS, had identified two independent dimensions of cognitive impairment in MS: frontal-executive dimension (attention, verbal fluency, recall), and posterior dimension (language and visuospatial functions). Connick et al. further report that language and visuospatial deficits account for 14% and 11% respectively of the total cognitive impairment in MS; and for 55% and 45% respectively of the variance not explained by fronto-executive impairments. Connick et al. have reported that language and visuospatial impairments formed independent deficits, a finding which was opposite from what we have observed in our study. By reporting isolated language dysfunction in both PPMS and SPMS groups, Connick et al. hypothesized that language and visuospatial impairments may be typical of cognitive dysfunction in progressive MS, where deficits due to randomly distributed focal white matter lesions are minimal and that the specific language and visuospatial deficits could in fact be attributed to mechanisms of neurodegeneration associated with progressive MS (Connick et al., 2013). So far this, together with the studies by Grossman (1995) and Mackenzie & Green (2009), have been the only studies that emphasised the need to assess language functions in pwMS and even suggested including language assessments into routine clinical assessments of people with progressive MS.

We believe that the reason behind different findings and therefore different conclusions between our study and that of Connick et al. was different methodologies employed to study cognition and different statistical techniques used for inferential statistics. In their study Connick et al. had employed the factor analysis on the ACE-R data, therefore it could be considered that, even though the aim was to investigate the dimensionality of cognitive impairment in MS, with that statistical technique, they were actually measuring the dimensionality of the instrument itself (ACE-R) and not

the dimensionality of the patient performance on it. Factor analysis is a common technique in psychometric studies used to evaluate the instruments to unravel how the individual items within the instrument group together. Since Connick et al. had performed a factor analysis on the ACE-R component items, therefore they had analysed what underlying functions the ACE-R measured. And indeed, the 'frontal executive domain' as they called it, or attention, learning and recall items, do account for 58% of the ACE-R test, and the language and visuospatial items comprise 42% of ACE-R (26 points and 16 points of 100 on ACE-R respectively). It could be argued that by using a factor analysis as a statistical technique Connick et al. have shown only how much of the ACE-R battery is comprised of items assessing which cognitive functions, and that their results could not be used to make inferences about cognitive profiles of MS patients. I believe that if replicated on healthy controls, the factor analysis would yield the exact same results, therefore their findings can only be used to make inferences about the dimensionality of the instrument, and not the study cohort. I believe that the dimensionality analysis presented in Chapter Five of this thesis could be considered a more suitable approach than a factor analysis to study cognitive profiles of MS patients, and that our findings in the language domain could be considered a more true representation of the dimensionality and cognitive profiles of pwMS than those from the Connick et al.'s study using the factor analysis. Therefore we did not support Connick et al's reasoning that language and visuospatial impairments could be indicative of neurodegeneration processes or characteristic to progressive MS.

III. Visuoperceptual functions

In our study we have shown the visuoperceptual impairments to be more common than language impairments among pwMS. Overall, 20% of pwMS at phase I and 55% of patients at phase II had some deficits with visuoperceptual performance, with or without additional deficits in other domains. This indicated that overall with disease duration more patients had accumulated deficits in the visuoperceptual domain.

In our study we found that, differently from the language where the deficits in that domain were indicative of a more cognitive global decline, in our sample the pwMS have also exhibited *isolated* impairments in the visuoperceptual domain. At phase I 7% of pwMS had isolated visuoperceptual impairment, compared to only 3% who had isolated visuoperceptual impairment at phase II. This

indicated that in some cases visuoperceptual problems may represent an initial symptom in the beginning of MS that gradually with disease duration spread into other domains. This would be in line with the pattern of MS symptoms observed in the clinical practice, where a portion of patients experience visual perception symptoms early in their MS. In our study we have shown that in MS visuoperceptual deficits can occur individually and even as the first symptom, as well as accompany other cognitive symptoms in a more global cognitive decline.

At phase I 16% of the SPMS patients, 18% of RRMS and 4% of PPMS patients failed the visuoperceptual cognitive domain. These individuals remained impaired on the visuoperceptual tests at phase II, while accumulating additional deficits in other cognitive domains. The three patients who had failed visuoperceptual tests at baseline, but not at follow-up were RRMS patients, and it could be speculated that their phase I performance may have been indicative of a relapse that was not present again during re-testing at follow-up. At phase II 22% of SPMS, 25% of RRMS and 23% of PPMS patients had failed the visuoperceptual domain, which was indicative that more pwMS developed visuoperceptual deficits with disease duration. Even though from analysing phase I data it seemed that the visuoperceptual impairments were more prevalent in the RRMS and SPMS, at phase II it was found that these deficits were equally distributed among all subtypes of MS.

Our findings at phase I were in line with the previous research that investigated visuoperceptual deficits comparing different types of MS. However, we have failed to reproduce the same pattern as described elsewhere at phase II.

Gaudino and colleagues (2001) have demonstrated similar findings to our phase I findings in their study, although they had used tests that measured visuoperception as well as visuospatial memory (therefore less pure reflections of visuoperception function, as ‘contaminated’ by involvement of memory functions). According to Gaudino et al., the specificity of visual perception and acquisition deficits to the RRMS and SPMS subtypes highlights the possibility of an increased susceptibility to visuospatial deficits with a disease course involving relapses and remissions. This was recently shown again by Planche et al. (2016) where the differences between RRMS/SPMS and PPMS remained even after controlling for age, gender, disability level, disease duration, and education level. Gaudino et al. further reasoned, referencing the work by Fischer et al. (1998), that since SPMS is

marked by initial RRMS disease course followed by progression, therefore the underlying disease pathology could be considered the same in these two MS subtypes, perhaps with specificity for decline in visuoperceptual recognition and learning functions. Despite these differences in the initial acquisition of visuospatial information, in their study disease course did not differentially affect the rate of forgetting, with no significant differences noted with regard to recall of newly learned information (Gaudino et al., 2001). They concluded that RRMS and SPMS are more prone to deficits when initially processing visuoperceptual information, and that their visuoperceptual problems could be the reason for memory problems, as the patients who exhibit deficits when initially processing visual information have more problems remembering it. However, we have only managed to observe these patterns at phase I, but not at phase II, therefore the generalizability of the above reasoning by Gaudino et al. should be considered with caution.

Moreover, when comparing our results to those of Gaudino et al., it is important to consider the different methodologies employed. In our study we have selected the visuoperceptual tests that captured a much wider array of visual cognition functions than those in Gaudino et al.'s battery. Our battery included tests assessing visual stimuli identification (recognition of what the object is), visuospatial functions (apprehension of spatial relationships among the objects) and analysis of other visual characteristics of the objects (size, length, orientation). Our visuoperceptual assessment was more thorough and included a larger domain of visuoperceptual functions, which can be concluded that the findings presented by Gaudino et al. can be generalizable only to visuospatial perception and memory. Therefore an alternative interpretation should be considered in order to model other visuoperceptual functions, especially later on in disease progression, as has been shown with our phase II results.

IV. Memory

In this study the two types of memory, verbal and visuospatial, were studied separately. We found verbal memory deficits to be more common, with 33% of pwMS failing verbal memory domain at phase I, and 34% at phase II. With regard to MS subtype, at phase I 27% of PPMS, 27% of SPMS and 20% of RRMS failed the verbal memory domain; and at phase II 19% of PPMS, 32% of SPMS and 22% of RRMS failed the verbal memory domain. As it can be seen, the SPMS patients have

consistently exhibited more frequent verbal memory deficits. Moreover, even though at phase I PPMS had similar frequencies of verbal memory impairments, at phase II SPMS patients were noticeably more commonly impaired than the other MS subtypes. It could be hypothesized that the reason behind these findings was that in our sample PPMS patients have had MS for a shorter amount of time, therefore had participated in fewer assessments and thus were less familiar with cognitive tests during the baseline assessment. This reasoning could potentially explain why the PPMS patients had overall, as a group shown higher frequencies of verbal memory impairment at baseline than at follow-up.

Now regarding the visuospatial memory, on these tests the patients had substantially lower prevalence of impairment in our sample. Only 5% of pwMS had failed the visuospatial memory domain at phase I, and 10% at phase II. It can be seen that even though the prevalence of impairment on visuospatial memory was much lower than that of verbal memory, it had a similar tendency where with time more patients had acquired visuospatial memory deficits. With regard to MS subtypes, at phase I 4% of PPMS, 5% of SPMS, 2% of RRMS failed the visuospatial memory tests, and at phase II 8% of PPMS, 14% of SPMS and 2% of RRMS failed the visuospatial memory tests. It can be noted that even though earlier in their MS (at phase I) the three groups of patients seemed to exhibit similar frequencies of visuospatial memory deficits, with disease duration (at phase II) the SPMS showed a tendency to develop more impairments in the visuospatial memory domain. This observation was in line with the previous finding regarding verbal memory, and it can be concluded, that even though some proportion of all MS patients fail these tests, the patients with the SPMS course were the most prone to memory problems (both verbal or visuospatial).

Our findings have partially replicated those reported by Gaudino and colleagues (2001) who have looked into group differences on verbal and visuospatial memory in different subtypes of MS using the same tests as in our study (CVLT and SPART). Their study showed qualitative differences between all three MS subtypes: PPMS and SPMS had more problems with verbal new learning than RRMS patients, and SPMS and RRMS patients had more difficulties with visuospatial new learning than PPMS patients (Gaudino et al., 2001). We have shown that among all subtypes SPMS had the most pronounced verbal and visuospatial memory deficits, however, we have shown limited support

for their finding regarding differences between PPMS and RRMS. In our data PPMS had more impairments on verbal memory only at phase I but not at phase II; and we have shown that not only the RRMS were not more likely to show more visuospatial memory deficits than PPMS at neither baseline nor follow-up assessments. Indeed, we have shown that the opposite was true in our data – the progressive MS types were more likely to show deficits on visuospatial memory tests than RRMS patients.

Gaudino et al. explain their findings regarding group differences among people with different types of MS in that PPMS and SPMS result in significantly greater deficits with regard to acquisition of new verbal information, with the SPMS showing a significantly higher failure rate (Gaudino et al., 2001). Even though we have replicated the higher frequencies of impairments in the verbal memory domain, and have associated that with SPMS, however, in our study we did not observe any differences between verbal learning and verbal recall at neither of the phases. Therefore our memory data does not support such interpretation and limits the generalizability of Gaudino et al.'s findings.

An alternative interpretation by Litvan et al. (1988) suggest that patients with MS may suffer from multiple memory impairment that occur at different stages of a theoretical memory system. PwMS are reported to have trouble processing information at the level of the articulatory loop, in retrieving verbal information from long-term storage and in rapidly processing verbal information. In summary, Litvan et al. state that pwMS have memory deficits primarily because of a deficit in a component of the working memory system, and slowed information processing speed. This is consistent with previous work noting a relation between memory performance and information processing abilities in MS (DeLuca et al., 1994).

The mechanism behind the memory problems in pwMS is further explained from a neuroanatomical perspective. Since knowledge representation is stored in gray matter, then patients whose primary CNS pathologic condition is in white matter should be expected to demonstrate cognitive deficits in components on an information-processing system that are responsible for transmitting information from the gray matter for activation or rehearsal purposes (Litvan et al., 1988). Litvan and colleagues further reason that if impaired articulatory loop was accepted as the reason behind memory deficits in pwMS, then controlling for encoding by forcing pwMS to encode deeply verbal information, or

by involving semantics, should aid in their long-term recall. They further elaborate that pwMS have been shown to make fewer mistakes and have higher accuracy of performance once they are given more time to process the information (Litvan et al., 1988). This again strengthens the assumption that memory deficits are inseparable from information processing speed deficits in pwMS.

V. Information processing speed

16% of patients in our sample had failed the information processing speed domain at phase I, compared to 29% of patients who failed at phase II. Overall in our sample there was a tendency for more patients to acquire information processing speed deficits with progression of their MS.

A split analysis of the different subgroups of MS at phase I has shown that 8% of PPMS, 24% of SPMS and 4% of RRMS patients failed the information processing speed domain. However, at phase II 19% of PPMS, 35% of SPMS and 10% of RRMS patients had exhibited deficits. It can be noted that information processing deficits occur in all groups of pwMS and increase in their frequency over time, and with higher predisposition towards the SPMS group. The differences between MS types were the highest for information processing speed in our study, indicating that this deficit was most notably associated with SPMS course, more so than the impairments on other cognitive domains.

This finding was in line with previous research that showed that general slowing occurred in all MS subtypes with higher slowing rates for the SPMS and PPMS subtypes, and RRMS patients seem to hold a position halfway between the controls and the PPMS/SPMS group (De Sonneville et al., 2002).

Not only have we observed that SPMS patients exhibit processing speed impairments more frequently, but in our study we also noted that their deficits have been more severe than those of PPMS or RRMS patients. This was not in line with a finding described by Foong et al. (2000) who reported little difference in the proportion of PPMS and SPMS patients who had severe cognitive impairment in their sample. The results by Foong and colleagues could be considered limited due to their small sample size (12 PPMS and 13 SPMS patients), while in our study the sample size was double of theirs for PPMS and triple for SPMS (50 RRMS, 26 PPMS and 37 SPMS). Therefore based on our findings the profiles of performance on information processing speed tasks should be considered different for SPMS and PPMS patients.

In MS information processing speed so far has been most extensively studied by De Sonneville and colleagues in their 2002 paper. In their study they argue that slowed information processing speed in MS is a result of not a single deficit, but of multiple deficits. To begin with, De Sonneville et al. argue that the information processing speed deficits exhibited by pwMS revolve not only around slowed times of giving a response, but also that pwMS tend to demonstrate slowed response speed compared to controls when they simply have to detect the mere presence of a stimulus. Moreover, when higher cognitive demands were imposed in the more complex tasks, differences in speed between controls and MS patients tend to increase dramatically. MS patients were more susceptible to distraction than controls: speed of performance deteriorated disproportionately in the presence of distractors. MS patients slowed down much more after making an error that controls, the extent of this delay suggesting an interruption of the ongoing process rather than an adequate adjustment, underscoring the existence of executive control weaknesses (De Sonneville et al., 2002). Therefore as shown by De Sonneville and colleagues, there are many reasons for poorer performance on the information processing speed tasks in pwMS.

As MS is known to cause problems with fine motor control and motor slowing as well as psychomotor slowing, the dissociation between their effects on slowed performance on cognitive tests has been the focus of several previous studies. Such dissociation is particularly important as SPMS patients tend to exhibit higher levels of both, and in order to successfully measure each of them separately, adequate measurement tools need to be employed.

Inordinate increase of reaction time in pwMS with task complexity suggest a slowing of mental processing independent of motor slowing (Archibald & Fisk, 2000; Rao, Aubin-Faubert & Leo, 1989), but others have failed to find this interaction (Litvan et a., 1988). Such inconsistencies in findings might be explained by many factors such as differences in tasks, severity of the disease, disease course, disease duration, and cognitive status of MS patients, or different statistical approach (De Sonneville et al., 2002).

Moreover, the influence of physical disability on performance on information processing speed tasks can also manifest through impaired vision, slowing of eye movement, sensory motor slowing and hand dexterity problems (De Sonneville et al., 2002).

Besides neurological disability, previous research has suggested that fatigue also tends to be associated with poorer performance on speed of information processing tests and needs to be controlled for during cognitive assessments as well (Denney et al., 2004). So far this has been successfully addressed only in one study by employing not only a healthy control group, but a clinical control group as well. In their study rheumatoid arthritis (RA) was selected as a clinical control because it is an autoimmune disease resulting in physical disability (although with no known involvement of the CNS); with comparable symptoms of pain and fatigue that might impact performance on cognitive tests; and these patients often take similar types of medication as pwMS (e.g. anti-inflammatories; immunosupressants) with potential side effects upon cognitive performance (Denney et al., 2005). In Denney et al's study (2005) slowing in the speed of information processing was characteristic only to MS patients, and RA performed no differently from the controls on tests of processing speed. This could be used to conclude that once the confounding variables (pain, fatigue, physical disability, medication) are accounted for, the core processes underlying cognitive slowing can be attributed only to MS CNS pathology (Denney et al., 2005). Future studies should consider not only enrolling a healthy control group, but also a clinical control group, such as RA, in order to dissociate among the many factors that contribute to poorer performance on information processing speed tests by MS group.

In our study the De Sonneville et al's (2002) multi-causal model presented earlier could be an explanation of why so many MS patients have performed so poorly on the PASAT test, with a significant number of patients only collecting the minimal score of zero. As shown by the findings presented by De Sonneville and colleagues, too many time constraints make the task too difficult for the pwMS to engage in (2002), therefore by not being able to engage in the task those with processing speed deficits are doomed to fail the task before they can even begin performing it. Too many time constraints in the PASAT test (new number every three seconds), make the patients distracted by their own mistakes, following which they cannot engage again with performing the test. As we have seen in our practice, and as was explained by De Sonneville's work, even though the PASAT is considered to be the standard measurement of information processing speed in MS, it is in fact a really poor measurement, as the nature of MS deficits doesn't allow the patients to successfully engage in the task, resulting in the non-informative scores of zero. The only patients that can engage

and therefore attempt to perform it are those that don't have the information processing speed deficits. Therefore by administering the PASAT on pwMS with information processing speed impairments their ability does not get successfully measured, and all that administering this test does is it only confronts the patients with their limitations and gets them frustrated and upset about the level of their abilities. Thus the PASAT could be argued to only be deemed suitable to measure attentional deficits on a limited sample of MS patients – those who are early in their MS and have good numerical ability (higher education levels). Other patients, because of their deficits, cannot successfully engage in performing this test. Limited application sample makes the use of PASAT in intervention studies limited, and more appropriate measurements of information processing speed should be considered.

Hence keeping all of the considerations regarding multiple causes of poorer performance on the information processing speed tests (De Sonneville et al., 2002) in mind, an appropriate test of information processing speed should have only the stimulus exposure time pre-determined and allow patients to respond at their leisure. Preferably to facilitate the administration such test should also be not confounded by patient's physical disability, such as the ability to move the arms or give fast verbal responses. Moreover, to ensure adequate interpretation, the information processing speed tests should have no involvement of other cognitive processes, such as memory or executive functions; something that the commonly used SDMT and PASAT tests had been criticized for (Piccinin & Rabbitt, 1999). Such confounder-free assessment would be particularly important if we consider that deteriorating processing speed has the potential to be considered a biomarker of MS-related CNS disease (Hoffman et al., 2007). Therefore developing information processing speed tests that are free of all problems associated with SDMT and PASAT is particularly important.

A potential candidate test has been developed and used in research involving cognitive ageing (Penke et al., 2010; Penke et al., 2012) and neurological diseases such as Parkinson's disease (Shipley et al., 2002), and CNS malignancy (Scotland et al., 2012). The Visual Inspection Time (VIT) test involves patients making simple decisions about a stimulus presented on the computer screen, with the duration of stimulus exposure being varied by the experimenter. This way the time for how long the participants see the stimulus is controlled and since the respondents can respond at their leisure the responses are not confounded by the time it takes the patients to move their arms or give verbal

responses. Our own pilot data supports VIT feasibility in MS, with evidence of clear disease-specific signal (Pettit, 2015). We believe that the VIT test can act as a suitable candidate to be used instead of PASAT and SDMT in further research as it free of disability confounders associated with PASAT and SDMT, is a more suitable measure for patients that have attention deficits (aforementioned problems with PASAT), easy to explain and well tolerable in neurological samples (Pettit, 2015).

In this study we have also shown that processing speed is one of the main cognitive deficits and an indicator of MS progression. Further research is needed to make any conclusions regarding the causality, but our findings did not contradict those presented by De Soneville and colleagues (2002) who had previously discussed that disruption in information processing speed may be considered the underlying cause for other cognitive impairments. In their study De Sonneville et al. have shown that processing speed is substantially correlated with SPMS type, disease duration and disability, which they then explained by CNS pathology.

Moreover, if we assume that the information processing speed is the underlying deficit behind other deficits in other cognitive domains, then a test that reliably measures information processing speed could be considered the ideal candidate to be the potential output variable as an indicator of disease progression or as a measure of the efficacy of treatment. More research needs to be done in order to investigate how likely the information processing speed is to represent a biomarker for MS progression and the extent to which it is independent from other MS symptoms. However, based on our pilot study using the VIT test (Pettit, 2015), it can be considered likely.

Previous research into processing speed shows that once given unlimited time to perform the task, the pwMS were still slower, but this time did not differ in accuracy of processing from controls, suggesting the importance of using time strategies in planning everyday life and job activities to compensate for or alleviate MS-related speed handicaps (Demaree et al., 1999; De Sonneville et al., 2002). Seeing our results in the light of those presented by De Sonneville and colleagues we raise a question - could it be that by measuring other cognitive functions (such as the number of words retrieved on a verbal memory task or the number of words in a category listed on a verbal fluency task) we are actually measuring the outcome of deficits in concentration abilities, but not the domain-specific deficits themselves? Could it be that concentration and information processing speed, rather

than other deficits hinder performance on other cognitive domains? More work for dissociation among memory problems and problems with information processing speed and attention are needed to disentangle among the causes for cognitive deficits in MS. Future work might consider focusing on replicating our study but with patients performing all the tests twice – first with limited time and the second time with unlimited time. It would be worthwhile to study how different settings of time given for performance might influence the cognitive outcome. Such study would yield results that could help build successful compensation and cognitive rehabilitation strategies to compensate for MS-related cognitive limitations by using time strategies to cope with daily life requirements and job activities.

8.2.5. Causes for cognitive impairment in MS

Even though there seems to be consensus in the literature that the SPMS type is associated with more prominent cognitive impairment, however, no clear explanation for what causes the cognitive deficits in MS has been presented. There have been attempts to name the specific profile of MS cognitive impairment as subcortical disconnection dementia (Piras et al., 2003), with it's distinctive features of reduced information processing speed causing intellectual slowing, attentional problems, impairment in abstract reasoning, problem solving, and memory dysfunction. The causes for subcortical disconnection dementia condition have been attributable to the interruption of the neural connections among cortical associative areas as well as between cortical and subcortical structures as a consequence of demyelination and axonal degeneration (Piras et al., 2003).

The biological cause for the distinctive profile of subcortical disconnection dementia could be explained by myelin breakdown which slows neural conduction along an axon and therefore slowing problem solving speed (Peters, 1996) and that nerves with thinner myelin sheaths are slower and less accurate in signal processing (Miller, 1994). In clinical practice this is considered to be represented by MRI abnormalities and the size and number of lesions (Thompson et al., 1990).

However, finding evidence to support the reasoning behind subcortical disconnection dementia theory is not so easy as previous studies have shown that increased MRI lesions do not correlate with the clinical course of the disease and cognitive deficit evolution (Piras et al., 2003; Foong et al.,

2000). This was well illustrated by Dominik Meier (Brigham & Women's Hospital, Boston), based on a time-series analysis of 24 brain MRI scans taken in a man with RRMS over 12 months. Dominik Meier has shown that throughout a span of one year numerous WMH grow and shrink, as if the disease is active, however, according to the clinical measures, the patient was stable during that year. This observation has been referred to as the so-called clinic-radiological paradox - abnormal spots on MRI often don't manifest in physical or cognitive symptoms (Barkhof, 2002), and physicians cannot always trace symptoms to a particular spot on a scan (Chen, 2012). This is further illustrated in a systematic review by Mollison et al. (under review) where correlations as low as r = - 0.30 between cognitive performance and white matter lesion burden have been reported.

Moreover, previous research indicate that the severity of cognitive impairment cannot be fully explained by the extent of abnormalities detected on conventional MRI images, and that other pathological abnormalities such as normal-appearing white matter are likely to be involved (Foong et al., 2000). Interestingly Meier et al. (2007) have shown that the extent of brain white matter lesions were not only not predicting functional degeneration, but also found that smaller lesions appeared disproportionally more damaging than larger lesions, with lesions in progressive MS smaller and of shorter activity than in RRMS. It has been further hypothesized that nonfocal, diffuse changes in the MS brain, especially axonal loss and mitochondrial dysfunction, prove better correlates of disability than total lesion load and have been associated with disease progression (Stadelmann, 2011). A neurodegenerative component of pathology is increasingly recognised in all forms of MS, but may be particularly salient to the progressive phase (Stadelmann et al., 2011; Stadelmann, 2011).

Thus, cognitive dysfunction could be related to disease peculiarity and not the time course and there could be several causes behind cognitive deficits (Piras et al., 2003). Although inflammation and neurodegeneration co-exist throughout the disease, they assume shifting prominence dependent upon the clinical stage, with neurodegeneration assuming greater pathogenic significance in patients with progressive disease (Reynolds et al., 2011). The mechanistic relationship between these components of the disease biology remains debated (Stadelmann et al., 2011); however, the relative lack of focal inflammatory white matter disease in PPMS compared to SPMS suggests that differences observed between these MS subtypes could be explained by contribution of multifocal inflammation (Connick et al., 2013). It could be argued that the poorer cognitive performance of the SPMS patients observed

in this study could be associated with the complex interactions behind both inflammatory and degenerative processes particular to this subtype of MS.

8.3. Evaluation of study methodology

As part of this research project not only I've attempted to answer the three research questions (dimensionality, predictors and perception of change), but I also tried to address the methodological problems that exist in many longitudinal studies. These methodological caveats tend to play a major role in representability of the results and in comparability of the findings among different longitudinal studies. In this section I address two of methodological aspects that I tried to ensure throughout my project – test battery optimization and recruitment (sample representativeness).

8.3.1. Test battery optimization

The phase II battery optimization procedure has served several purposes in this study. To begin with, the optimization procedure has helped shorten the follow-up assessment while ensuring that the test battery coverage remains intact. Moreover, since novel fields of interest have emerged in the time span from the development of the phase I battery to the implementation of the follow-up assessment, it was necessary to incorporate the additional questionnaires to capture the new data of interest. This is especially important when working with clinical samples such as MS where individuals tire quickly (thus it's very expensive to administer time-consuming assessments that are redundant or not sensitive), and where individuals are highly heterogeneous in their symptoms (thus a wide range of domains needs to be evaluated).

To begin with, pragmatic considerations identified through feedback from phase I participants on the tolerability of the assessment schedule indicated that the phase I assessment was too long and that an unchanged assessment schedule was likely to impact negatively on recruitment to phase II. In order for the results of the longitudinal study to be reliable I needed to do everything in my power to ensure high recruitment rates. And indeed, making the phase II test battery twice shorter by removing redundant or insensitive items has undeniably increased the follow-up recruitment rates and the overall patient satisfaction with the study. Especially since pwMS tend to tire quickly, at this stage an assessment of over three hours would have been unbearable and would have resulted in lower

recruitment rates. It is also important to note, however, that the changes made did not affect any other test parameters, such as the test setting and sequence, and that all attempts were undertaken to ensure that the results from the phase I and phase II would be comparable.

It is important to note that to the best of my knowledge this longitudinal study so far has been the first to optimize its assessment battery to adjust for the participants' feedback. Most longitudinal studies in MS have a few tests and low coverage, but even studies with a wide selection of tests and high coverage (such as Morrow et al., 2009; Strober et al., 2014) didn't report critically evaluating their tools.

I believe that the attempt to optimize the phase II battery has led to higher follow-up recruitment rates in this study than those reported by other longitudinal studies in MS. I believe that this has been a strong methodological aspect which has highly benefited the representativeness of my follow-up sample and that this practice should be considered to be employed in future longitudinal studies. In addition, I would strongly suggest other researchers to aim to optimize their test batteries prior to each of the follow-up assessments to make their assessments still comparable to the previous ones, while responding to the advances of the latest developments in the field.

8.3.2. Sample representativeness

This section includes a discussion of follow-up recruitment rates, and the characterization and representativeness of the study cohort.

This study had high recruitment rates (over 70% for both patients and controls), which can be considered high in comparison to the follow-up rates of other longitudinal studies in MS (Table 1.2., Chapter One). Not only this study had high recruitment rates to be deemed representable of the patient sample, but it also had high follow-up rates and was representable of the control sample. Few longitudinal studies in MS have recruited controls, and if they did, their follow-up rates haven't been presented in the write-up. I consider it to be important to discuss not only the patients', but the controls' follow-up recruitment rates as well, as not having a representative sample yields results that are difficult to generalize and interpret.

Moreover, this study allowed the participants to have the assessments administered at the convenience of their own homes, which in turn has helped secure higher recruitment rates by involving the individuals with severe neurological disability, and also those with full-time employment or other responsibilities (such as inflexible schedule, lack of transportation or having nowhere to leave the children) to be able to take part in research. Without the researcher having to come to their house for assessment, these individuals would not have been able to participate, resulting in lower representativeness of the follow-up sample.

One major disadvantage of the longitudinal study design is that we can only generalize the findings of the individuals who come back for follow-up assessments, with the data of those who don't come back being lost. There are several reasons why participants do not come back for follow-up assessments. To begin with, they may be feeling too ill or have their neurological disability progressed to the point where it's difficult to move. Besides the physical reasons, the individuals who don't come back may have high levels of depression, more cognitive disability, etc., which, due to higher drop-out rates, remain uncaptured in many longitudinal studies. Therefore it could be considered that the studies with lower follow-up recruitment rates are indeed studying the longitudinal changes of those individuals who have lower levels of disability or depression, just because such samples are more convenient to approach, making them easier to study. Only by ensuring high recruitment rates I can be certain that the findings of this study can be generalizable to all pwMS, not only those who are least affected, highly motivated, and have the time and means to make their way to the hospital buildings for clinical visits.

Moreover, these arguments imply that the longitudinal studies which do not allow the re-assessments to be performed at patient homes, underestimate the true progression of MS symptoms. This brings me to the next issue – the generalizability of findings from longitudinal studies in MS. If other studies don't perform actions to accommodate for patients who have progressed (optimizing the battery and allowing the assessments to be administered at patient homes) then those studies are limited to including only the patients with less severe courses of MS. This could imply that such studies indeed monitor not the progression of MS, but instead the life of the 'survivors', i.e. of the individuals with the most benevolent courses of MS who have their lives least affected. Indeed, providing with

research conditions tolerable only for the relatively healthy patients results in research findings applicable only to a subsample, instead of all of the patients with MS.

Accommodating for the individuals whose MS has severely progressed and making the research settings more tolerable for them was one of the strong methodological aspects of this study, allowing to provide with results more indicative of the actual situation in MS progression than previous longitudinal studies. This was an achievement; however, even with doing so, in this study the follow-up data from about one in four participants was still not captured. This indicates that in the future there is potential for more to be done in order to increase the appeal of participation in the follow-up assessments for the participants, and to ensure better representativeness of the dataset and generalizability of the findings.

8.3.3. Clinical applications of study findings

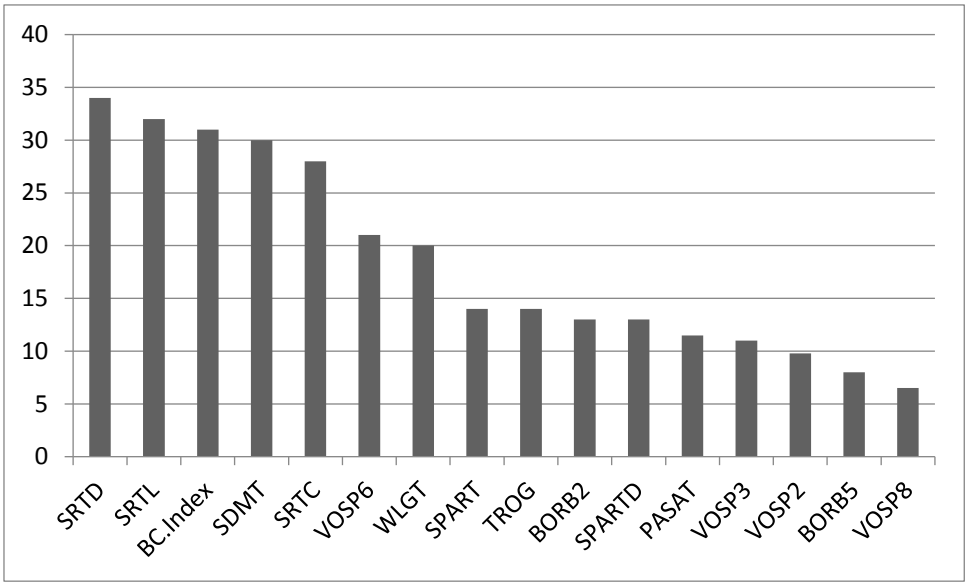
This section was allocated for the clinical application of this study's findings. Here I identified the most sensitive and specific cognitive tests for administration for pwMS during regular neuropsychological assessments, and for tracking symptom progression.

To begin with, from analysing the trajectory of change I've shown that both pwMS and healthy controls have exhibited changes in performance on the cognitive tests, the direction of which not always being negative. Upon temporal evaluation both controls and pwMS showed increased and decreased performances, and the trajectory and extent of those changes differed depending on a cognitive test. Taking that into account, the tests that can be considered most suitable for picking up MS-related deficits should be sensitive (those that pwMS fail at baseline but even more at follow-up) and specific (controls don't fail them). On those tests the pwMS should exhibit only one trajectory of change, representing development of new impairments attributable only to the progression of MS.

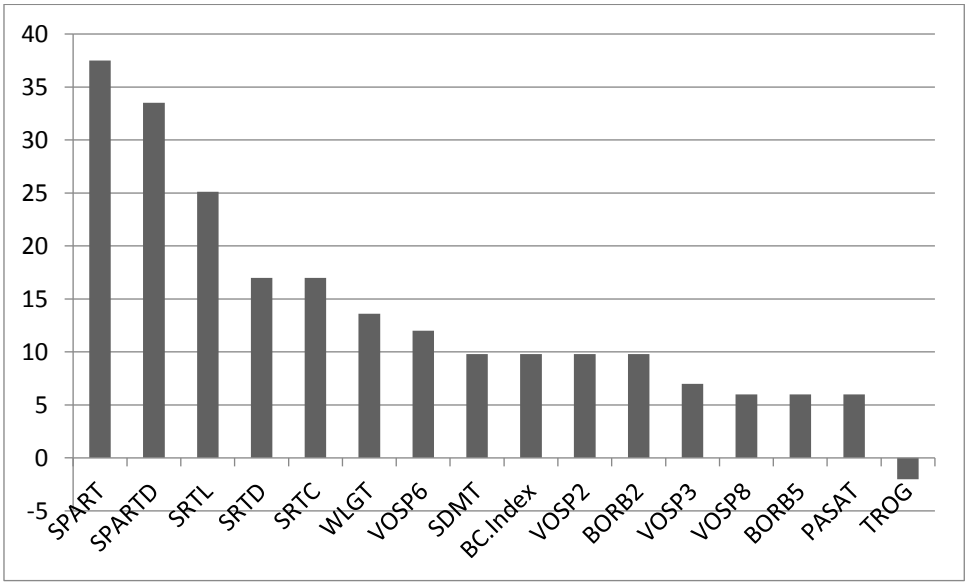
The suitability of the tests for use in clinical practice was investigated by inspecting the true frequencies of pwMS failing each cognitive task (percentage of pwMS minus percentage of controls). The tests, arranged in the order from the most sensitive and specific to the least at phases I and II can be seen in Figure 8.1.

Figure 8.1. True frequencies (%) of pwMS failing each cognitive task at phases I and II

a) True frequencies at phase I



b) True frequencies for phase II



Note. True frequencies were calculated by subtracting the percentage of controls from the percentage of pwMS that failed the test

Abbreviations: SRT – Selective Reminding Test (SRTL– Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

As it can be seen from Figure 8.1., different items were the most sensitive and specific to MS-related cognitive impairment at phases I and II. At phase I the items of verbal memory, spontaneous speech, and visual information processing speed have topped the list. However at phase II, the items of visuospatial memory, and only then followed by verbal memory, were highest on the list. Moreover, the conclusion about which tests should be advised to be used in everyday clinical practice is even more difficult, as some of the items that were quite sensitive at phase I, had lost their sensitivity at phase II (such as TROG and PASAT). Therefore it can be concluded that, even though cognitive impairment in MS should be regarded as a multidimensional construct and for thorough assessments tests measuring visuoperceptual and language abilities should be included as well, for quick screening at everyday clinical practice the BRBN battery (without the PASAT) can be considered to be the most suitable instrument. In this study the PASAT has been shown to be a problematic test, with its instructions being too difficult to comprehend not only to the pwMS but to the control participants as well.

Based on the results of this study, administering the whole BRBN battery should be regarded as a sufficient examination of neuropsychological changes in pwMS. However, even though proposed by some authors, narrowing down the examination to administration only of SDMT as a sole component of BRBN is not enough, as administrating only the SDMT (especially as has been shown in Figure 8.1. b) would not suffice in picking up the MS-related deficits.

Moreover, part of the reason why the SPART test was shown to be so sensitive to impairment at phase II could be because performing it involves hand movement and hand-eye coordination, functions that tend to deteriorate together with MS progression. It could be suggested that an approach is needed to make the SPART test more applicable to be administered on individuals with upper limb paralysis, as in this study three patients were unable to perform the task due to inability to move their hands; although their visuospatial memory was considered to be unaffected.

In this study the 2SD cut-off was employed and, due to the reference group that was used, the cut-off values produced for the majority of BRBN tasks were higher than those proposed by the test authors. Therefore even though these cut-off values were more representative of our reference control group, as a consequence these tests became more sensitive (and less specific) to detect cognitive

problems in the MS sample. It could possibly be that the suitability of the BRBN battery tests to detect MS-related cognitive impairment observed in this study might have been inflated by the way the cut-off values were calculated.

8.3.4. Definition of impairment

This section of the discussion centres around the complex issue of definition of cognitive impairment and how I believe that it was successfully dealt with in this study. This aspect is important as by being thorough in how one defines impairment results in the reliability of the findings, and the clinical application of the results. In this study the impairments were defined separately for each test item and for each phase of the study. Moreover, specific definition for cognitive domain failure was created to address the dimensionality of cognitive impairments. I believe that the approaches employed in this study were successful and resulted in greater validity of the study findings.

One of the advantages of this study was that separate cut-off scores for impairment definition were created for the phase I and phase II assessments. These adjustments produced differing cut-off values for the majority of the cognitive tests (eleven of sixteen), and although small differences between cut-offs for most of the items, for some items the differences were quite significant (e.g. 20-point (out of 60) increase in PASAT cut-off). I believe that by adjusting the cut-offs I managed to avoid systematic differences in test administration, practice effect, and healthy cognitive ageing. To my knowledge no other longitudinal study up to date has re-adjusted their cut-offs in order to redefine the normal range of performance. This is very important in future longitudinal research, since, as has been shown by this study, both patients and healthy controls tend to vary in their performance.

This study has attempted to capture that normal fluctuation, and make attempts to incorporate that variation into the result analysis in order not to over-estimate the variation in performance exhibited by the patient participants. This attempt as illustrated in this study should encourage other longitudinal researchers to consider external factors not related to MS, such as practice effect, healthy cognitive ageing and systematic differences in test administration, prior to running their result analyses; as accounting for these external sources of variation could potentially help explain the heterogeneity of findings from the longitudinal studies into cognitive changes in MS.

In this study an approach was chosen to define impairment on a domain when a patient failed at least two tests within a domain. The motivation behind this approach was that it was less sensitive than failing at least one test, but definitely more specific, while allowing comparable application across all tested domains. I believe that this approach has led to valid findings especially since the domains have an unequal number of cognitive tests in them. Moreover, this approach has also minimized the chance of picking up false positives, where the patients might have failed one test in a domain because they didn't understand the task or due to fatigue. Being more conservative and avoiding such false positives is also an important aspect since this study could be considered experimental in that some of the tests employed here (majority of visuoperceptual tests and language tests) have never before been used in MS. Therefore I believe that by using this approach I have managed to capture the domain-specificity of MS impairment while maintaining a good balance on the trade-off between sensitivity and specificity, therefore avoiding both the type I and type II errors.

8.3.5. Reference control group

One of the strongest aspects of this project was that not only I managed to collect a representative patient sample, but that I also managed to collect and follow-up a corresponding control group. This group of participants was well-matched to the patients in terms of age, gender and education, and in majority was comprised of patients family members and friends, thus they were comparable in terms of their background and cognitive reserve to the patient group. I managed to ensure that my control group differed from the patient group only in one aspect – they didn't have MS. Moreover, the follow-up recruitment rates were the same for patient and control samples therefore it can be concluded that there had been no systematic differences between individuals showing up for follow-up between these two groups. Therefore it can be safe to infer that all the differences found in longitudinal changes in cognition between these two groups could be attributable to factors caused solely by MS.

By having a reference control group I managed to account for the effects such as normal fluctuation in performance, cognitive ageing, practice effect, and systematic differences in administration between the phases. To the best of my knowledge, no longitudinal study of cognitive changes in MS up to date has accounted for all of those effects. Therefore for the aforementioned reasons future

longitudinal studies in MS should ensure that matched control samples should followed as well, and that their recruitment is controlled with the same rigour as that of the patients.

8.3.6. Other strengths and weaknesses of the work

In our study we have demonstrated a finding which could be of interest to other researchers planning to perform repeated assessments on pwMS, either through an observational or interventional longitudinal study design. The weaknesses of this work would be associated with some of the tests chosen, and the strengths of this work would be associated with the methodology for the analysis of longitudinal change.

Even though language deficits are rare in MS and were not expected to be picked up at a frequent level, however, it could be argued that other language tests could have been better candidates for including into the assessment battery rather than the BCT that had been selected for this study. As it was noted in Chapter Six, the patients have exhibited no change on the BCT, but the controls have exhibited a dramatic 7% drop in their performance on the BCT from phase I to phase II. This discrepancy was inconsistent with the profile of deterioration in cognitive performance shown on other tests where both controls and patients have deteriorated (visuoperceptual items). On those items the patients have deteriorated to a similar or to a higher extent than the controls, and therefore on those items that decrease in performance was linked to the effects of cognitive ageing or development of age-related eye problems. However, with regard to the BCT, only the controls have deteriorated, but not the patients, and since both groups had been matched for age, we searched for an alternative explanation behind this observation.

One candidate for a potential explanation of why the controls but not the patients had performed worse at follow-up than at baseline could be related to the nature of the BCT administration. While other cognitive tests were highly structured and during administration were presented in a question-and-answer format, the BCT administration could be considered the most unstructured of all of the tasks performed. In this task the participants are asked to describe the activity taking place in the stimulus picture. No other guidelines are given – the participants respond at their leisure and decide themselves the contents, the length and the style of their descriptions. Therefore the way that the task

is presented allows each participant to understand its requirements in their own way. This had resulted in a vast range of the number of words written by each participant, some of them producing one sentence descriptions with others producing a full page of text with hundreds of words. Currently this task has been shown by previous literature as a suitable candidate to pick up language problems in an individual patient, however, in order for the BCT to be used in the future for cross-sectional comparisons more structure should be introduced into this task allowing for more comparable results between individuals.

However, thinking in retrospect, having considered this limitation of the BCT, it is still unclear whether there could be a more structured way to measure spontaneous speech. In our study the language domain had good coverage, and was comprised of tests measuring both language comprehension and language production. Future researchers might wish to consider alternative ways how to measure spontaneous language production in a more structured manner that would be less affected by individual interpretations and that would allow comparison to an individual's performance on other, more structured tests.

Other potential sources for limitations in a longitudinal project could be centred around practice effect and controlling for it. We believe, that by using the methodology employed in this study we have managed to avoid those limitations, and turn them into the strengths of this study. The method for quantification and individual analysis of longitudinal change in performance employed in this study centred around using the healthy controls as a reference point. The reasoning behind that was that this way the items with the most practice effect could be identified by analysing the changes in performance in the control group, and then controlled for by incorporating average control change into the analysis of the average patient change for each item. This way we have controlled for not only the practice effect, but also for the effects of cognitive ageing and administrative differences if there had been any.

For example, there were several items where the controls had exhibited practice effect. The controls showed better performance at phase II than at phase I on visuospatial memory (SPART), information processing speed (SDMT, PASAT, WLGT), and naming (VOSP2) tasks. This was not a surprising finding as practice improves performance on memory tasks, even though in this study we aimed to

control for that by employing alternative versions of the BRNB battery tests. Moreover, having done the naming task for the second time, many participants had reported remembering doing it the first time and managed to recognise several of the stimulus materials from three years ago. The increased performance of the control group on the information processing speed task indicated that by performing it for the second time they already knew how to perform it, and spent less time trying to understand the requirements of the task, or maybe even remembered performance strategies acquired during the phase I assessment.

The patients, however, had demonstrated an increase (though slightly smaller) in performance on the verbal fluency, visuospatial learning (but a decrease on visuospatial retrieval), no change on the naming task, and deterioration on the information processing speed tasks (SDMT and PASAT). Only by incorporating the average change in the control group we have managed to show that the 'no change' on the naming task actually indicated emergence of mild deficits in the patient group, as, compared to the controls who have improved on this task, the patients had performed worse. While the age matched controls have shown that it is normal to remember some of the stimuli from three years ago during re-assessment, the patients as a group have failed to exhibit this pattern. Therefore if we hadn't incorporated the average change of controls into our methodology of estimating change, we would have overlooked that emerging deficit and treated it as unchanged performance.

This illustration was particularly evident for the information processing speed tests in our study. While the patients had exhibited mild deterioration on these items (1% decrease on SDMT and 10% decrease on PASAT), the controls showed improved performance (4% increase in SDMT and 5% increase in PASAT), indicative of practice effect. Therefore by incorporating the average change of the controls into our methodology, we have adequately inflated the deterioration exhibited by patients (from 1% decrease to 5% decrease on SDMT and from 10% decrease to 15% decrease on PASAT). This is particularly important, as since the SDMT and PASAT are the most commonly used tests in research, it could be considered that other studies which did not employ this methodology and did not account for practice effect exhibited by controls, tended to underestimate the actual deterioration in patient performance on the information processing speed tasks. Although it is a well established finding that information processing speed tends to deteriorate in MS, it could be considered that the

actual extent of deterioration reported by previous studies could have been underappreciated. Again this stresses the importance of employing reference control groups in longitudinal studies in order to correctly estimate the amount of cognitive change, especially on the often studied information processing speed domain.

The results of our 3-year follow-up study showed that on some of the tasks the control participants had a tendency to perform better at follow-up, therefore indicating that those patients who had exhibited non-different performance on those tests, had actually deteriorated, even though that deterioration was masked as lack of improvement. These findings are in line with those of another 2-year follow-up which showed that PPMS and SPMS subtypes were associated with lack of improvement compared to controls and RRMS on the PASAT and SDMT, but not on other tasks of the BRBN (Huijbregts et al., 2006). They also conclude that future longitudinal studies should consider that lack of improvement with practice in progressive MS on the SDMT and PASAT indicates a short-term manifestation of cognitive deterioration (Huijbregts et al., 2006), rather than no change.

8.4. Chapter summary and study conclusions

This project has been the first representative attempt to study domain-specificity, predictors, and perception of longitudinal change in cognitive impairment in MS. No MS study up to today has had such high coverage of functions assessed and overall sample representativeness. The richness of this study data and high follow-up percent, allowed us to be confident of the study findings and to ensure their generalizability. With this project not only we've provided an answer about the domain specificity of MS impairment, but also related it to a rich dataset of other measures of interest – demographic, clinical, and cognitive reserve variables. In addition, not only we've investigated the actual change in performance, but we've also analysed the patients' own perspective into their longitudinal change, which has never been done before in MS in this way.

The results regarding the domain-specificity of MS-related cognitive performance have been discussed extensively with regard to MS subtype and other causes behind cognitive symptom

development. The strengths and weaknesses of the work have been addressed, and suggestions for future research and clinical practice have been made.

We believe that the questioning of methodologies behind the data collection and analysis was one of the strongest aspects of this project. By doing so we've addressed the issue of why different longitudinal studies in MS have yielded different results, and contributed to the better understanding of such a heterogeneous condition. By attempting to be methodologically precise we believe that we have managed to show evidence of controlling for such external factors as natural variation in performance, cognitive ageing, practice effect, and administration differences, by utilizing a very representative control group, and accounting for their variance when calculating the standardized scores of change.

Due to the aforementioned reasons, this study could be considered to be a useful resource to researchers who aim to investigate the progression of cognitive deficits in MS longitudinally. We believe that this study could be of interest not only to those who specialize in cognitive functioning in MS, but also to those who question the methods employed in clinical research to define impairments and to account for individual differences.

9. Appendices

Appendix A. Normative values of phase I tests

Table 1. Published cut-off values for each test used to define cognitive impairment. MS-specific cut-off norms are indicated in **bold**

	Test	1.5SD cut-off	2SD cut-off	Publication
BRBN	SRTL	* < 29years 38; 30-39years 33; 40-49 years 32; 50-59years 33; 60-69years 23; 70-79years 17. **** 32** ***** <39years 37, 40-49years 33, 50-59years 25, >60years 11**	* <29years 34; 30-39years 27; 40-49 years 25; 50-59years 28; 60-69years 16; 70-79years 11. **** 26** ***** <39years 32, 40-49years 28, 50-59years 19, >60years 3**	*(Larrabee, Trahan, & Levin, 2000) ****(Boringa et al., 2001)** *****(Rao, 1990)**
	SRTC	* <29years 30; 30-39years 21; 40-49 years 19; 50-59years 22; 60-69years 13; 70-79years 9 **** 16**	* <29years 25; 30-39years 13; 40-49 years 12; 50-59years 16; 60-69years 6; 70-79years 2 **** 9**	*(Larrabee, et al., 2000) ****(Boringa et al., 2001)**
	SRTD	*18-29years 11; 30-39years 8; 40-49 years 9; 50-59years 9; 60-69years 6; 70-79years 6 **** 6**	*18-29years 10; 30-39years 7; 40-49 years 8; 50-59years 9; 60-69years 5; 70-79years 4 **** 5**	*(Larrabee et al., 2000) ****(Boringa et al., 2001)**
	SDMT	*<12 YoE 57, >13YoE 62 **** 38** ***** 49**	*<12 YoE 56, >13YoE 62 **** 32** ***** 44**	*(Sheridan et al., 2006) ****(Boringa et al., 2001)** *****(Parmenter, Weinstock-Guttman, Garg, Munschauer, & Benedict, 2007)**
	PASAT	*** <12YoE 34; >13YoE 36** **** 33** *****31**	*** <12YoE 29; >13YoE 32** **** 28** *****26**	***(Rao, 1990)** ****(Boringa et al., 2001)** *****(Parmenter, et al., 2007)**
	SPART	**16**	**14**	**(Boringa et al., 2001)**
	10/36 SPARTD	**5**	**4**	**(Boringa et al., 2001)**
	WLGT (60s)	*<59 years 14, >60 years 10 **11	*<59 years 12, >60 years 8 **8	*(Tombaugh, Kozak, & Rees, 1999)

				**(Gladsjo et al., 1999)
VOSP	VOSP2	**15	*<50 years 16, >50 years 15; ** 13	*(Warrington & James, 1991) **(Bonello, Rapport, & Millis, 1997)
	VOSP3	**15	*<50 years 16, >50 years 15; ** 14	
	VOSP5	**9	* 8 ** 9	
	VOSP6	**18	* 18 ** 17	
	VOSP7	**8	*7 ** 7	
	VOSP8	**9	* 6 ** 8	
BORB	BORB2	NA	NA	
	BORB3	NA	NA	
	BORB4	NA	NA	
	BORB5	NA	NA	
	TROG	Bishop: 15 of 20 items passed		(Bishop, 1982)
	BC.Index (Sum of picture variables)/number of words	Bschor reports that healthy individuals typically come up with an average of 88.2 (SD=56.9) words, and on average 20.2 (SD=8) picture variables.		(Bschor et al., 2001)
	PPT	22	21	(Hulst, 2012)
	KDT	22	21	(Hulst, 2012)
	TTT	22	21	(Hulst, 2012)
	GNT	15	13	(Warrington, 1997)
	MPNWDS	46	46	(Rewaj, 2013)
	Orchard test	9	9	(Hulst, 2012)
	SPMS	27	26	(Hulst, 2012)

Abbreviations: pwMS – people with MS, YoE – years of education, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

Appendix B. Normality of distributions of phase I variables

Table 1. Exploration of normality of distributions of phase I variables

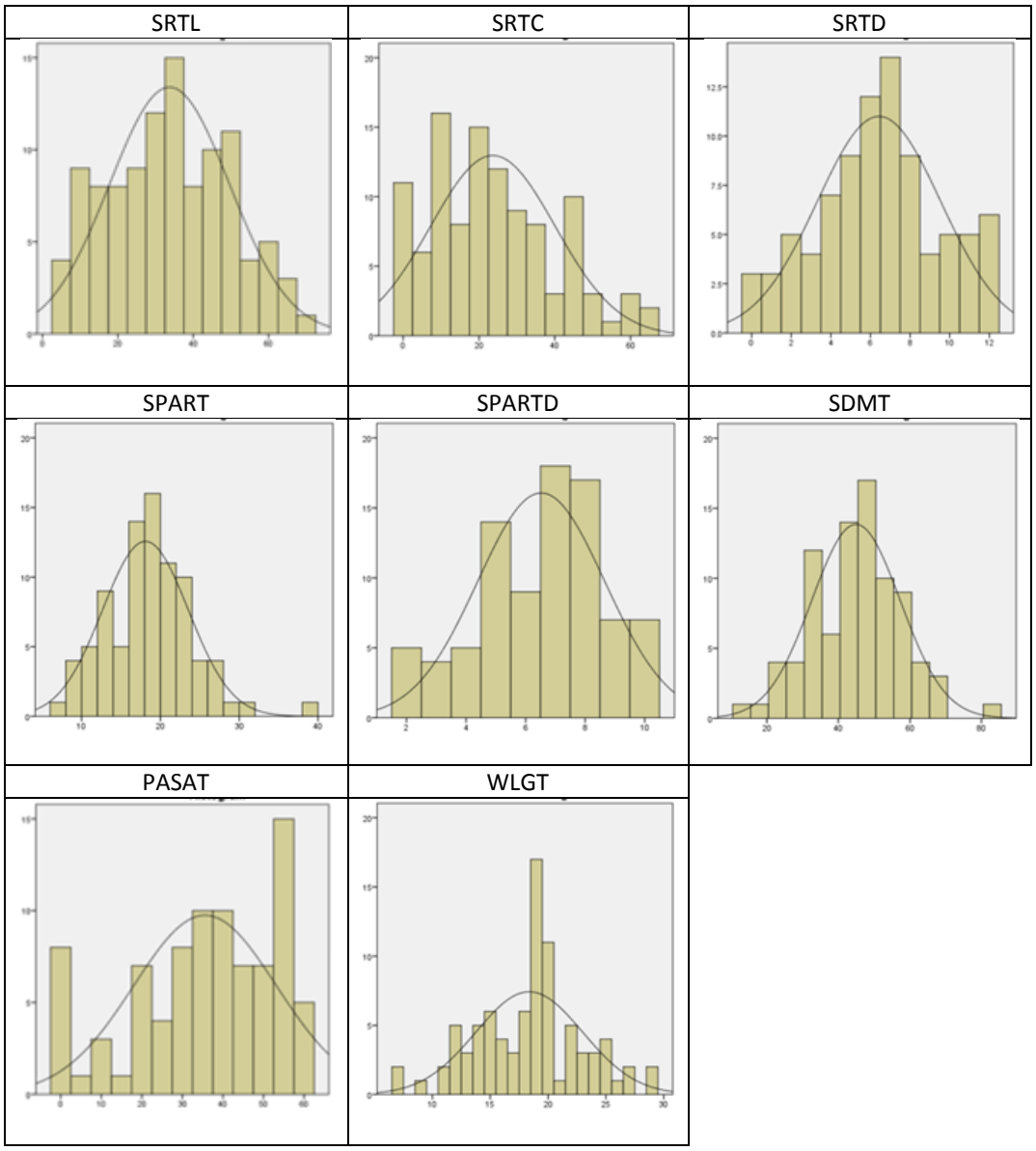
	PwMS (n=108)			Controls (n=33)		
Test	**N**	**W**	**p**	**N**	**W**	**p**
SRTL	107	0.978	0.071	29	0.922	0.182
SRTC	107	0.954	0.001	29	0.949	0.469
SRTD	107	0.970	0.044	29	0.744	0.001
SDMT	106	0.992	0.867	28	0.940	0.348
PASAT	105	0.924	<0.001	29	0.763	0.001
SPART	105	0.969	0.039	29	0.941	0.360
SPARTD	105	0.949	0.002	29	0.948	0.459
WLGT	108	0.980	0.216	32	0.906	0.099
VOSP2	105	0.950	0.002	31	0.894	0.064
VOSP3	105	0.845	<0.001	31	0.888	0.052
VOSP5	105	0.326	<0.001	29	0.273	<0.001
VOSP6	105	0.583	<0.001	31	0.405	<0.001
VOSP7	105	0.618	<0.001	31	0.591	<0.001
VOSP8	105	0.410	<0.001	30	0.398	<0.001
BORB2	103	0.939	0.001	30	0.960	0.666
BORB3	103	0.923	<0.001	30	0.949	0.467
BORB4	102	0.858	<0.001	30	0.920	0.166
BORB5	102	0.881	<0.001	30	0.878	0.037
TROG	101	0.872	<0.001	29	0.788	0.002
BC.Index	90	0.992	0.886	27	0.976	0.922
PPT	104	0.903	<0.001	28	0.797	0.003
KDT	104	0.901	<0.001	28	0.782	0.002
TTT	103	0.769	<0.001	28	0.762	0.001
GNT	101	0.881	<0.001	29	0.829	0.007
MPNWDS	100	0.598	<0.001	27	0.794	0.002
Orchard test	97	0.410	<0.001	25	0.405	<0.001
SPMT	102	0.848	<0.001	25	0.815	0.004

Note. Explorations of normality were performed with Shapiro-Wilk test

Abbreviations: PwMS – people with MS, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task, GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

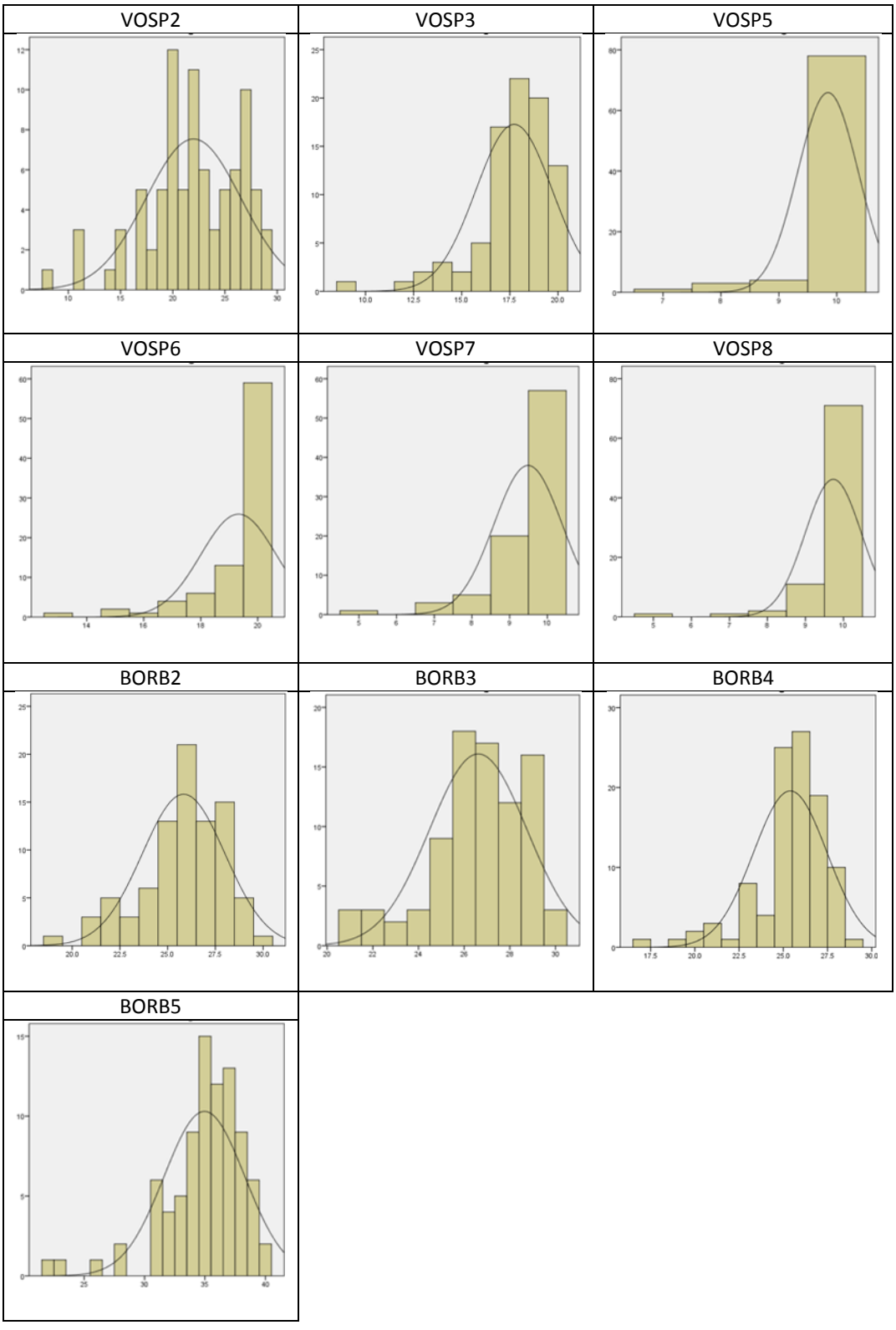
Figure 1. Histograms of distribution of *pwMS'* performance scores at phase I on individual test variables with a superimposed normative curve

1. BRBN items: SRTL, SRTC, SRTD, SPART, SPARTD, SDMT, PASAT, WLGT. Phase I patients (n = 108)



Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task

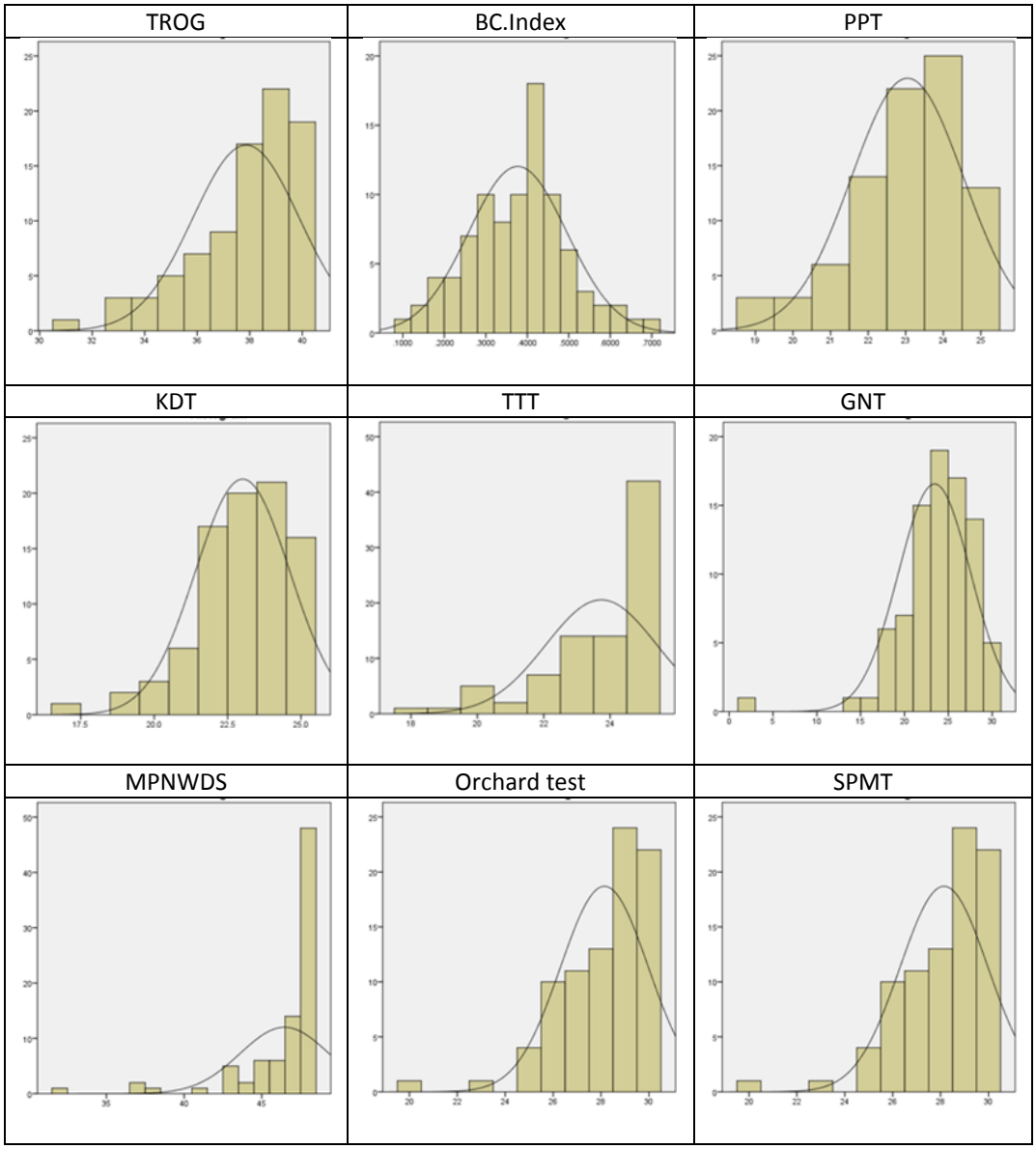
2. Visuoperceptual items: VOSP2, VOSP3, VOSP5, VOSP6, VOSP7, VOSP8, BORB2, BORB3, BORB4, BORB5. Phase I patients (n = 108)



Abbreviations: VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task,

VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task

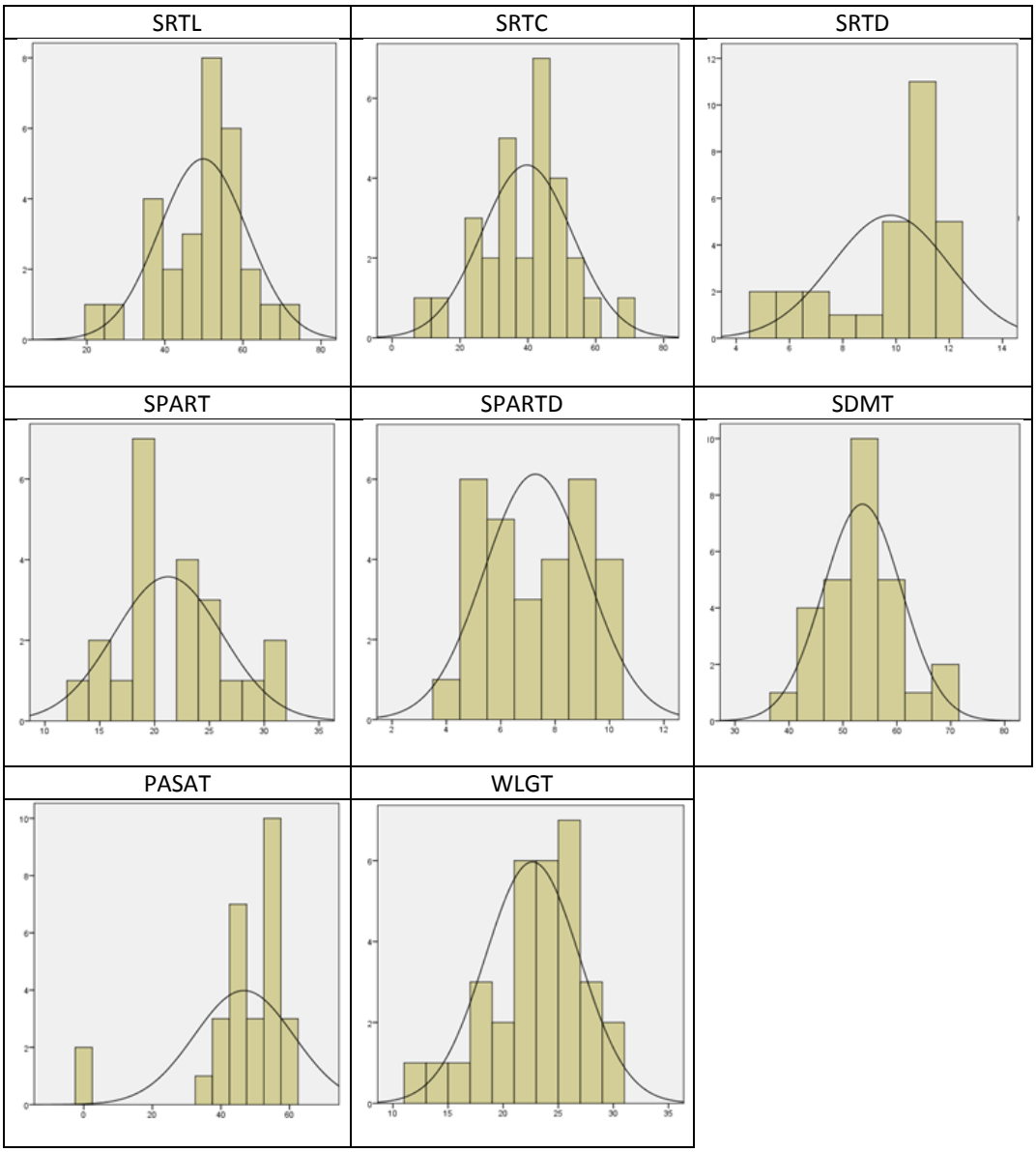
3. Language items: TROG, BC.Index, PPT, KDT, TTT, GNT, MPNWDS, Orchard test, SPMT. Phase I patients (n = 108)



Abbreviations: GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

Figure 2. Histograms of distribution of *controls'* performance scores at phase I on individual test variables with a superimposed normative curve

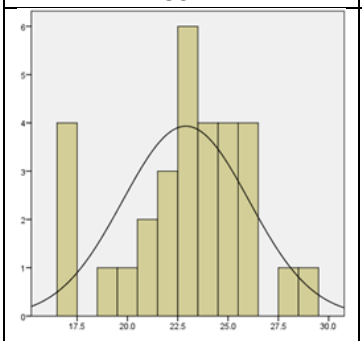
1. BRBN items: SRTL, SRTC, SRTD, SPART, SPARTD, SDMT, PASAT, WLGT. Phase I controls (n = 29)



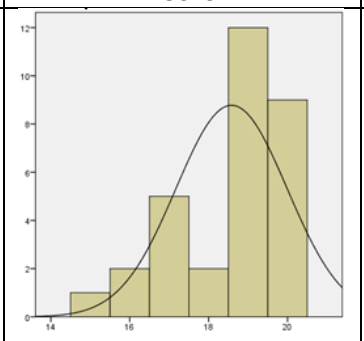
Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART – items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task

2. Visuoperceptual items: VOSP2, VOSP3, VOSP5, VOSP6, VOSP7, VOSP8, BORB2, BORB3, BORB4, BORB5. Phase I controls (n = 29)

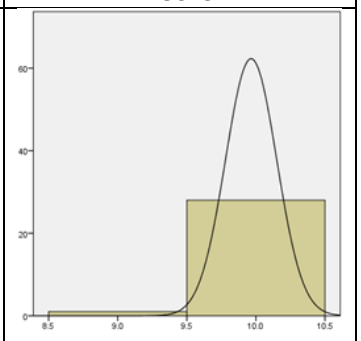
VOSP2



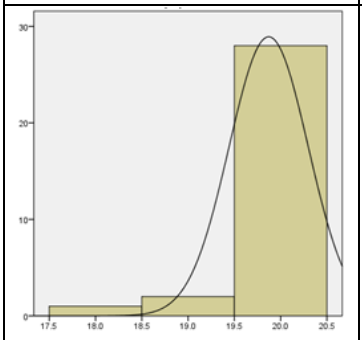
VOSP3



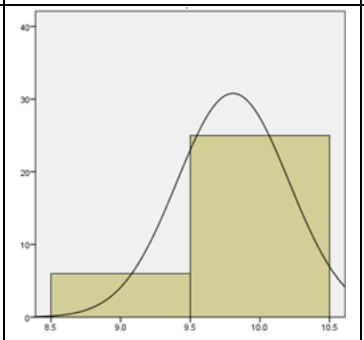
VOSP5



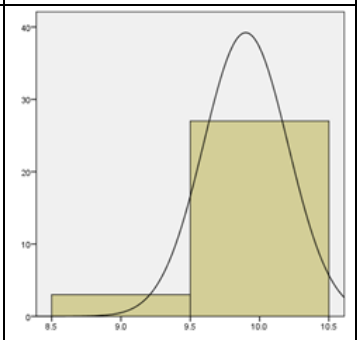
VOSP6



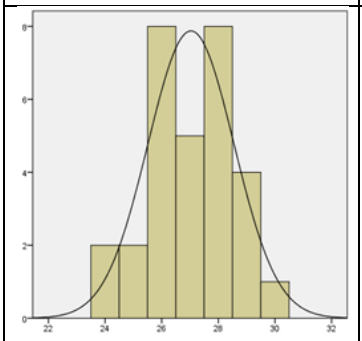
VOSP7



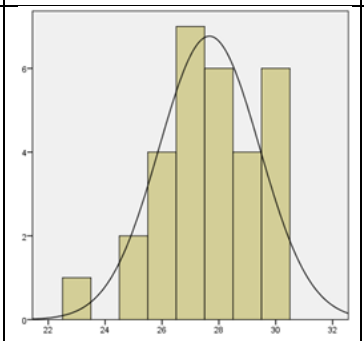
VOSP8



BORB2

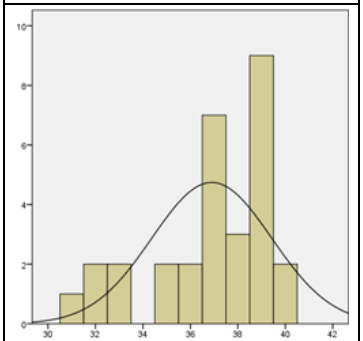


BORB3



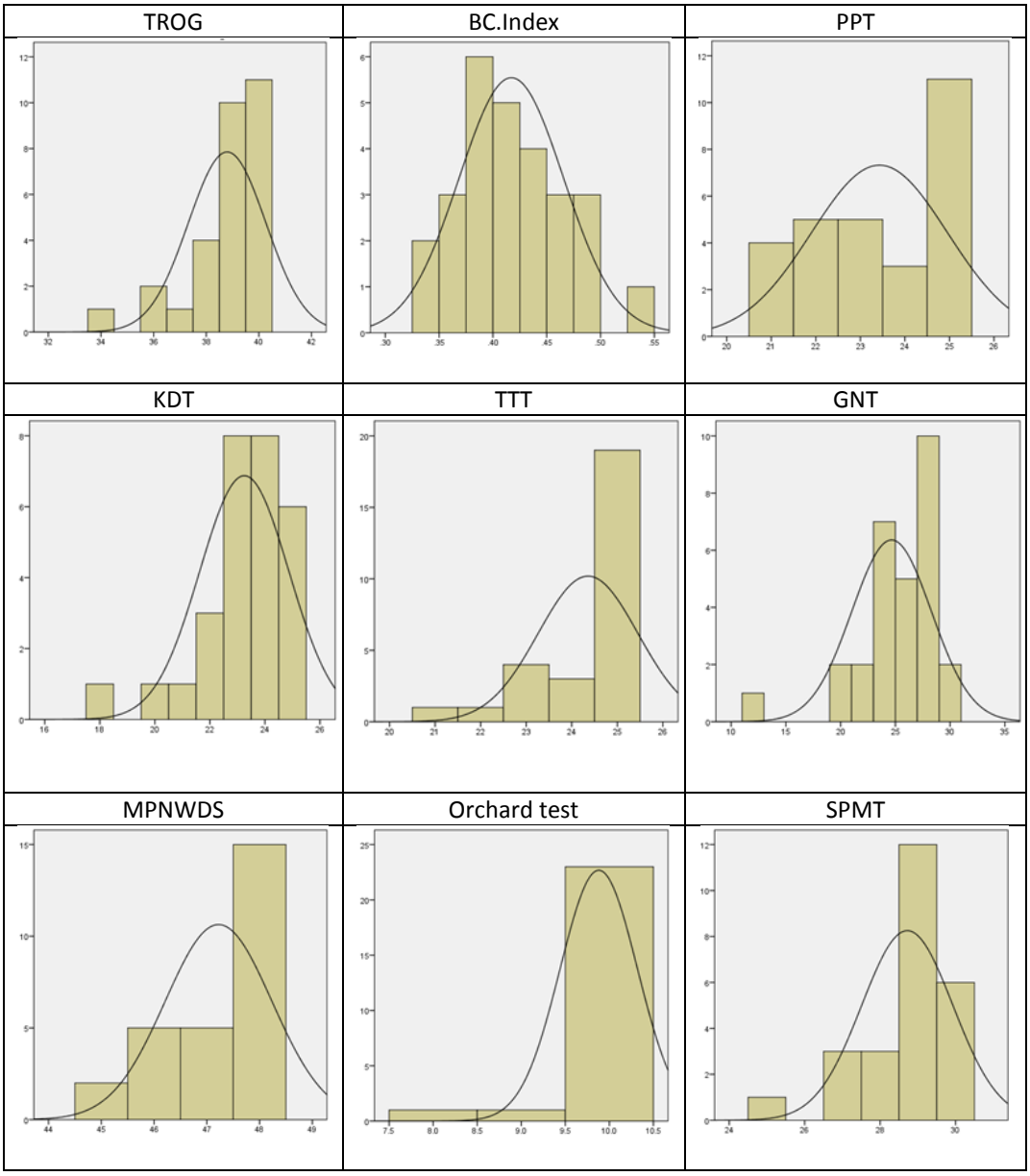
BORB4

BORB5



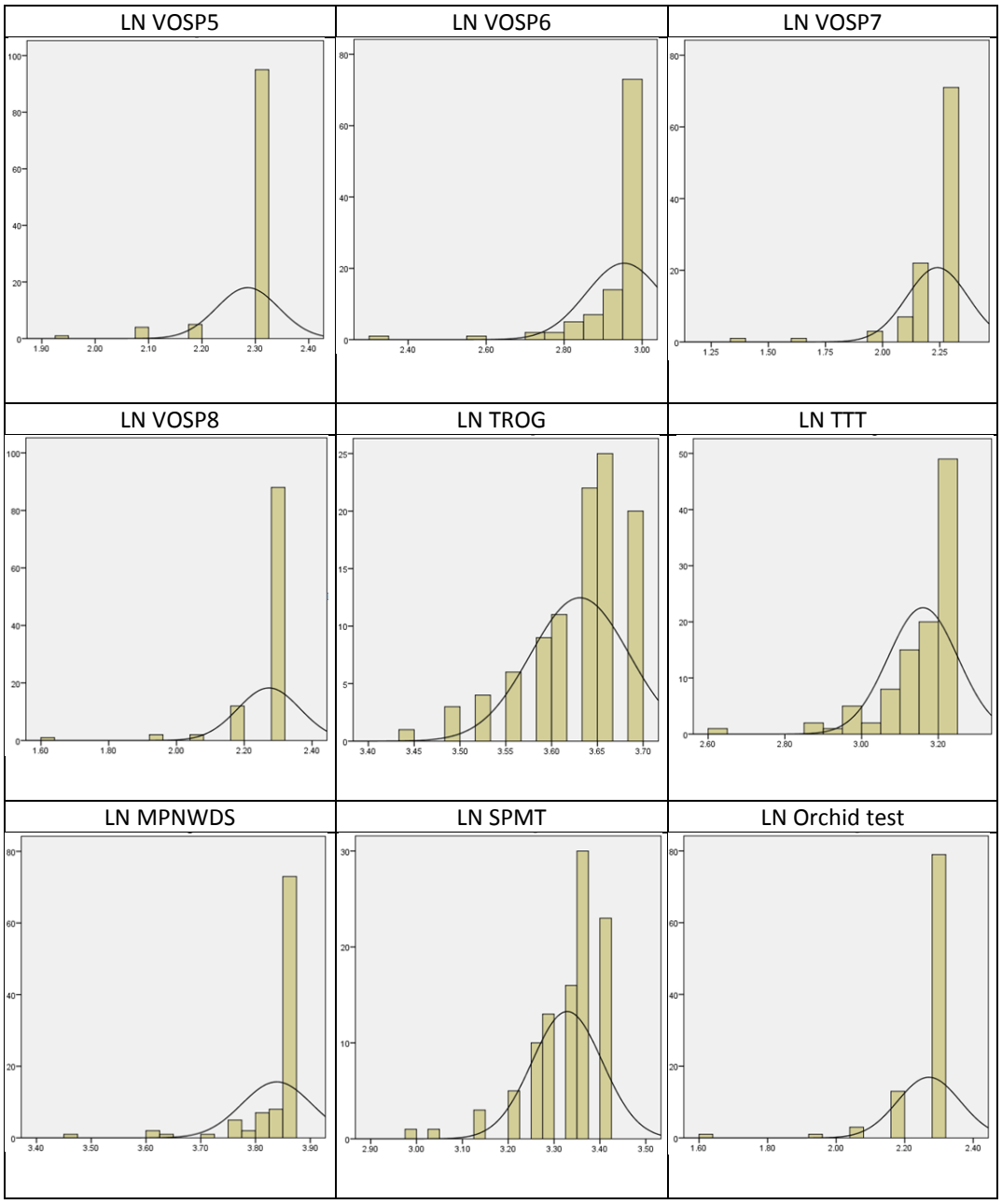
Abbreviations: VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB3 – Size Matching Task, BORB4 – Line Orientation Matching Task, BORB5 – Position of Gap Matching Task

3. Language items: TROG, BC.Index, PPT, KDT, TTT, GNT, MPNWDS, Orchard test, SPMT. Phase I controls (n = 29)



Abbreviations: GNT – Graded Naming Test, TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, PPT – Pyramids and Palm Trees Test, KDT – Kissing and Dancing Test, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), SPMT – Sound – Picture Matching Task.

Figure 3. Histograms of logarithmically transformed VOSP5, VOSP6, VOSP7, VOSP8, TROG, TTT, MPNWDS, SPMT and Orchard test variables. Phase I patients (n = 108)



Abbreviations: LN – natural logarithmic transformation, VOSP – Visual Object and Space Perception Battery (VOSP5 - Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task), TROG – Test of Reception of Grammar, MPNWDS – Minimal Pairs Non-Words Task, TTT – Tomato and Tuna Test, SPMT – Sound – Picture Matching Task.

Table 2. Properties of the distributions of the logarithmically transformed items. Phase I patients (n = 108)

Item	Skewness (SE)	Shapiro-Wilk Test
LN(VOSP5)	-3.851 (0.236)	W = 0.300, p < 0.001
LN(VOSP6)	-4.012 (0.236)	W = 0.536, p < 0.001
LN(VOSP7)	-3.870 (0.236)	W = 0.553, p < 0.001
LN(VOSP8)	-4.817 (0.236)	W = 0.346, p < 0.001
LN(TROG)	-1.199 (0.240)	W = 0.866, p < 0.001
LN(TTT)	-2.749 (0.238)	W = 0.664, p < 0.001
LN(MPNWDS)	-3.452 (0.241)	W = 0.561, p < 0.001
LN(SPMT)	-1.830 (0.239)	W = 0.816, p < 0.001
LN(Orchard test)	-4.826 (0.245)	W = 0.394, p < 0.001

Abbreviations: Visual Object and Space Perception Battery (VOSP5 – Dot Counting Test, VOSP6 – Position Discrimination Task, VOSP7 – Number Location Task, VOSP8 – Cube Counting Task,) TROG – Test of Reception of Grammar, TTT – Tomato and Tuna Test, MPNWDS – Minimal Pairs Non-Words Task, SPMT – Sound – Picture Matching Task.

Appendix C. Validity of phase II tests

Table 1. Evidence of validation of tests in the phase II battery

Test	Validation sample	Functions assessed, or sensitivity to lesions in brain areas	Validation study
EDSS	572 male MS patients	Pyramidal, Cerebellar, Brainstem, Sensory, Bowel and bladder, Visual and Cerebral (mental) functions	(Kurtzke, 1970)
ACE-R	241 subjects (AD 67, FTD 55, LBD 20; mild MCI 36; controls 63	Dementia screening	(Mioshi, et al., 2006)
SRT	Right (n = 20) and left (n = 22) temporal lobe epilepsy patients and controls (n = 49)	Verbal memory and in particular wordspan, correlated best with left temporal lobe abnormalities	(Bell, Fine, Dow, Seidenberg, & Hermann, 2005)
SPART	82 MS patients	Visuospatial memory and delayed recall , Correlates with parietal lesion load	(Lazeron et al., 2005)
SDMT	82 MS patients	Sustained attention and concentration Correlates with frontal, parietal and temporal lesion load	(Lazeron, et al., 2005)
PASAT	82 MS patients	Sustained and divided attention and information processing speed correlates with subcortical brain systems and white matter tract atrophy; Correlates with frontal, parietal and temporal lesion load	(Lazeron, et al., 2005)
WLGT	31 studies with 1,791 participants	Correlates moderately high with focal frontal (r=.54) and temporal (r=.61) injuries	(Henry & Crawford, 2004; Ross, 2003)
VOSP2	150 controls, and 55 right- and 52 left- hemisphere lesion cases	Object perception, Lesions in posterior right hemisphere	(Warrington & James, 1991)
VOSP3	150 controls, and 55 right- and 52 left- hemisphere lesion cases	Object perception , lesions in right hemisphere	(Warrington & James, 1991)
VOSP6	200 controls, and 74 right- and 75 left- hemisphere lesion cases	Space perception, Lesions in right hemisphere	(Warrington & James, 1991)
VOSP8	200 controls, and 74 right- and 75 left- hemisphere lesion cases	Space perception, Lesions in right hemisphere	(Warrington & James, 1991)

Test	Validation sample	Functions assessed, or sensitivity to lesions in brain areas	Validation study
BORB2 and BORB5		Right hemisphere lesion group is slightly worse at length location, and more noticeably worse at location discrimination than left-hemisphere lesioned patients. Results suggest general right-hemisphere specialization for visual processing.	(Riddoch & Humphreys, 1993)
TROG	70 controls	Grammatical comprehension	(Bishop, 1982)
Boston Cookie Theft Test (BCT)	34 patients with a degree of severity of aphasia ranging from slight to severe	Broca's Aphasia, posterior left prefrontal cortex, most notably Broca's area.	(Goodglass & Kaplan, 1983)
Cognitive reserve and cognitive leisure activities questionnaire	RR (n = 28), SP (n = 6), and PP (n = 2).	Cognitive reserve	(Sumowski et al., 2010)
BDI-II	Meta-analysis of 26 studies	Depression	(Beck, Steer, & Carbin, 1988)
MSIS-29	233 patients	Correlations with The 59 item Functional Assessment of Multiple Sclerosis (FAMS), the Medical Outcomes Study 36 item Short-Form Health Survey (SF-36), and the 12 item General Health Questionnaire (GHQ-12) provide evidence for the convergent and discriminant validity of MSIS-29 as a measure of the physical and psychological impact of MS.	(Riazi, Hobart, Lamping, Fitzpatrick, & Thompson, 2002)

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI-II – Beck's Depression Inventory 2nd Ed., MSIS-29 – 29 item Multiple Sclerosis Impact Scale, EDSS – Expanded Disability Status Scale, SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – delayed retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words), MS – multiple sclerosis, AD – Alzheimer's Disease, FTD – Fronto-Temporal Dementia, LBD – Dementia with Lewy bodies, MCI – Mild cognitive impairment.

Appendix D. Dimensionality of cognitive impairments of each individual

Table 1. Phase I ten subject-cluster solution. In this matrix the rows represent the 61 pwMS, and the columns represent individual cognitive tests.

0	Normal. Represents performance ***above*** the cut-off score, defined by 2SD below the Phase I (n=29) control mean
1	Impaired. Represents performance ***below*** the cut-off score, defined by 2SD below the Phase I (n=29) control mean

		Verbal memory			Visuosp. memory		Processing speed			Visuoperceptual						Language	
10-cluster	count	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
1	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
1	3	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	6	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1
1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	11	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0
1	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	15	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	17	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0
1	18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	19	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	20	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1
1	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	22	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1
1	23	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
1	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	25	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0
1	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

		Verbal memory			Visuosp. memory		Processing speed			Visuoperceptual						Language	
10-cluster	count	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	28	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1
1	29	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
1	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	31	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	33	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	34	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0
1	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	36	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
2	37	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0
3	38	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1
3	39	0	0	1	1	0	1	0	0	1	0	0	1	0	0	0	0
4	40	1	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0
4	41	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
4	42	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0
4	43	1	1	1	0	0	0	0	0	0	0	1	0	1	0	0	0
4	44	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4	45	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	1
4	46	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
5	47	1	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0
5	48	1	0	1	0	0	0	0	0	0	0	0	1	0	0	1	1
5	49	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1
5	50	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0
5	51	1	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0
6	52	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	1
7	53	1	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0
7	54	0	1	1	0	0	1	0	1	0	1	0	0	0	0	0	1
7	55	1	1	0	0	0	1	1	1	0	1	0	0	0	0	0	1
7	56	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0
8	57	1	0	1	0	0	1	1	1	0	0	1	0	0	0	1	1
8	58	0	0	0	0	0	1	1	1	0	0	1	0	0	1	1	0
9	59	1	0	1	0	1	0	0	1	0	1	0	0	1	1	1	0
9	60	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0
10	61	1	1	1	1	1	1	0	1	1	0	0	1	1	1	0	0

Table 2. Phase II two subject-cluster solution. In this matrix the rows represent the 61 pwMS, and the columns represent individual cognitive tests.

0	Normal. Represents performance ***above*** the cut-off score, defined by 2SD below the Phase II (n=23) control mean
1	Impaired. Represents performance ***below*** the cut-off score, defined by 2SD below the Phase II (n=23) control mean

		Verbal memory			Visuosp. memory		Processing speed			Visuoperceptual						Language	
10-cluster	count	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
1	3	0	0	0	1	1	0	0	0	1	0	0	1	0	1	1	0
1	4	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
1	5	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	6	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
1	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	9	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	11	0	0	0	0	0	1	1	0	0	0	1	0	1	1	1	0
1	12	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
1	13	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	14	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0
1	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	16	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
1	17	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0
1	18	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0
1	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	21	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	23	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	25	0	0	0	0	0	1	1	1	1	0	0	0	0	0	1	0
1	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	29	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0
1	30	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
1	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

		Verbal memory			Visuosp. memory		Processing speed			Visuoperceptual						Language	
2-cluster	count	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	32	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
1	33	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
1	34	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0
1	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	36	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
1	37	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0
1	38	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
1	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	40	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
1	41	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	1
1	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1	44	0	0	0	0	0	1	1	0	1	0	0	1	0	1	0	0
1	45	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	46	1	1	1	1	0	1	1	0	0	1	0	1	0	0	0	0
2	47	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0
2	48	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0
2	49	1	1	1	0	0	1	1	1	0	0	1	0	1	1	0	0
2	50	1	1	1	0	0	1	1	0	0	0	0	1	0	0	0	0
2	51	1	1	1	0	0	1	1	0	0	0	0	0	0	0	0	0
2	52	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
2	53	1	1	1	0	0	1	1	0	1	1	0	0	0	0	0	0
2	54	1	0	1	0	0	1	1	0	1	1	0	0	0	0	0	0
2	55	1	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
2	56	1	1	1	1	0	1	1	0	0	0	1	1	0	0	1	0
2	57	1	1	1	0	0	1	1	0	1	0	0	1	1	1	1	0
2	58	1	1	1	0	0	1	1	0	0	0	1	0	1	0	0	0
2	59	1	1	0	0	0	1	1	0	0	1	0	0	1	1	0	0
2	60	1	1	1	0	0	1	0	0	0	1	1	1	0	0	0	0
2	61	1	1	1	1	0	1	0	0	1	1	1	0	0	1	1	0

Table 3. Four subject-cluster solution for longitudinal change. The rows represent 61 pwMS and the columns represent individual cognitive tests.

-1 **Resolved impairment**. Represents cases that had performed ***below*** the cut-off at Phase I, but ***above*** the cut-off at Phase II. The cut-off scores were defined for the Phase I and Phase I assessments as performance below 2 SD below control mean.

0 **Unchanged**. Represents cases that performed identically (either above or below the cut-off) both at Phase I and Phase II.

1 **New impairment.** Represents cases that had performed ***above*** the cut-off at Phase I, but ***below*** the cut-off at Phase II.

		Verbal memory			Visuosp. memory		Processing speed			Visuoperceptual						Language	
4-cluster	count	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	1	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1
1	2	0	0	0	0	0	0	1	0	0	1	-1	1	0	0	0	0
1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	4	-1	0	-1	1	0	0	0	-1	1	-1	0	1	-1	0	0	0
1	5	1	0	1	0	0	0	0	0	0	0	-1	0	0	0	0	0
1	6	-1	-1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0
1	7	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0
1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
1	9	-1	-1	-1	0	-1	0	0	0	0	0	0	-1	0	0	0	0
1	10	1	0	-1	0	0	-1	0	0	-1	0	0	0	0	0	0	-1
1	11	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1
1	12	-1	0	-1	0	0	0	0	-1	0	0	0	0	1	1	0	-1
1	13	0	1	0	0	0	0	0	0	0	0	0	-1	0	1	-1	-1
1	14	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	15	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	-1
1	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	17	-1	-1	0	0	0	0	1	0	0	-1	-1	1	0	0	0	0
1	18	0	0	0	0	0	0	1	1	0	0	-1	0	0	0	0	0
1	19	0	0	0	0	0	1	1	0	0	1	-1	0	0	0	-1	0
1	20	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0
1	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	23	0	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	0
1	24	0	0	0	0	0	0	0	-1	0	0	-1	1	0	0	0	0
1	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	26	1	-1	-1	0	0	0	0	0	0	0	-1	0	0	0	-1	0
1	27	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0
1	28	0	0	0	0	0	1	1	0	0	-1	0	0	0	0	0	-1
1	29	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0

		Verbal memory			Visuosp. memory		Processing speed			Visuoperceptual						Language	
4-cluster	count	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
1	30	0	-1	0	0	-1	0	0	0	0	0	-1	0	0	0	0	-1
1	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
1	32	-1	0	0	0	0	0	0	1	1	0	-1	0	0	0	0	0
1	33	0	0	0	0	1	0	0	-1	0	0	0	0	0	0	0	-1
1	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
1	35	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	-1
1	36	1	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0
1	37	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0
1	38	1	0	0	0	0	0	1	-1	1	0	0	0	0	0	0	-1
1	39	0	0	0	0	0	1	1	-1	1	0	0	-1	0	0	0	-1
1	40	0	0	0	-1	1	0	0	0	-1	0	0	-1	0	0	0	0
1	41	0	-1	0	0	0	0	0	-1	0	-1	1	0	0	0	0	-1
1	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
1	43	0	0	0	-1	-1	1	0	0	0	0	0	0	0	0	0	-1
1	44	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0
1	45	0	0	-1	0	0	1	0	0	0	0	0	0	0	0	0	0
1	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	47	0	0	0	0	-1	1	1	0	0	0	0	-1	0	0	0	0
1	48	0	0	0	0	0	0	1	0	0	0	0	1	-1	0	-1	1
1	49	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0
1	50	0	0	-1	0	0	1	1	0	0	1	0	0	0	1	0	0
1	51	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	1
1	52	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0
1	53	-1	0	0	0	0	0	1	-1	1	0	0	1	0	1	-1	0
1	54	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	55	0	0	0	0	0	0	1	-1	0	1	1	0	-1	0	1	0
2	56	0	0	1	1	0	0	0	-1	0	0	1	1	0	0	1	0
2	57	1	1	0	1	0	0	0	0	1	1	0	0	0	1	1	0
3	58	1	1	1	0	0	1	1	0	0	0	1	0	0	0	0	0
3	59	1	1	1	0	0	1	1	0	0	0	0	-1	0	-1	0	-1
3	60	1	1	1	0	0	1	1	0	0	0	1	0	1	0	0	0
4	61	1	1	1	0	0	0	0	-1	1	0	-1	1	1	0	0	0

Appendix E. Dimensionality analyses with multiple imputation

1. Section Aim

The aim of the following insert was to rerun the Chapter Five dimensionality analyses (sections 5.3. and 5.4.), but this time with the data from the 21 pwMS with missing observations included. This supplementary analysis investigated the robustness of the dimensionality analyses conducted in Chapter Five, and it was also used to determine whether the Chapter Five results might have been biased towards the more healthy individuals (i.e. those without missing data).

2. Methods

2.1. Design

Just like the analyses in Chapter Five, this insert addressed the issue of disentangling whether the cognitive impairment in pwMS could be considered more of a unidimensional or multidimensional construct, and the longitudinal pattern of changes in domain-specificity of individual cognitive impairments.

In this supplementary insert the dimensionality analyses were carried out following the same methodology as presented in the methods section of Chapter Five. However, in this appendix, in addition to complete case analysis, subjects with missing observations were included as well.

2.2. Dealing with missing data

The cluster analysis method tends to assume the absence of missing data and is only able to include the observations for which every variable was measured. There are two ways how such analyses can be performed – as complete case analyses (excluding all participants with any missing observations) or by performing imputation to fill in the missing observations. Both methods of handling missing data have their strengths and weaknesses, therefore in order to address the aim of this thesis all cluster analyses have been performed twice – first as complete cases analyses (Chapter Five), and the second time with the missing values imputed (Appendix E).

2.2.1. Complete cases analysis

In the complete case analyses (Chapter Five) all cases with missing observations were excluded by performing list-wise deletion. If the missingness mechanism is missing completely at random (MCAR), a complete case analysis is sensible, although it may well not use all the available information in the data. However, in our study the cause of missingness was not MCAR since we were able to identify the source of missingness (Figures 2-4 in Results section). Therefore if the missingness mechanism is not MCAR, as was in our study, some authors argue that list-wise deletion might not be the most sensible approach (Carpenter & Kenward, 2007).

2.2.2. Imputation

The second way of dealing with missing observations is performing imputation to appropriately use all the information present in a dataset with missingness, therefore avoiding the biases, inefficiencies, and incorrect uncertainty estimates that can result from dropping all partially observed cases from the analysis. In social sciences missing data is most commonly dealt with ad-hoc methods of imputation, such as mean/mode/median imputation, however, they have been reported to lead to serious biases in variances and covariances (Honaker, King & Blackwell, 2015).

Another common way of dealing with missing data is Last Observation Carried Forward (LOCF), where for each patient their subsequent missing responses are set equal to their last observed response. However, this approach is not fit for studies where longitudinal change is expected as it decreases the sensitivity to detect the actual change (false negative). LOCF also does not help deal with missing observations at phase I, as it was the patient's first visit, and there are no previous values to carry forward. Moreover, the LOCF method has received much criticism, since it is neither valid under general assumptions nor based on statistical principles, and, according to Carpenter and Kenward, it is not a sensible method, and should not be used (2007). The group of patients who complete, but who share similar characteristics and responses to a specific patient prior to withdrawal, will usually give a better estimate of the distribution of the missing values than the last response before the patient withdrew (Carpenter & Kenward, 2007).

Multiple imputation (MI) is a general method that incorporates the uncertainty into the imputation process. One of the main reasons to use MI is the fact that all relevant data-collection information, both observed and unobserved, can be incorporated into the imputation. MI is comprised of three stages: imputation stage, in which the missing data are imputed; analysis stage, in which each complete data set is analysed using a complete-data technique; and the last stage, in which the results from the analysis are combined in order to yield a final result that combines the uncertainty in the data and the uncertainty due to missing values (Harel & Zhou, 2007). Multiple imputation has been shown to reduce bias and increase efficiency compared to list-wise deletion, ad-hoc methods of imputation, and LOCF methods (Honaker, King & Blackwell, 2015).

2.3. Multiple imputation

2.3.1. Imputation method

The R package 'Amelia II' performs MI to data with missing values by employing expectation-maximization with bootstrapping (EMB) algorithm (Figure 1) (Honaker & King, 2010; Honaker, King & Blackwell, 2015).

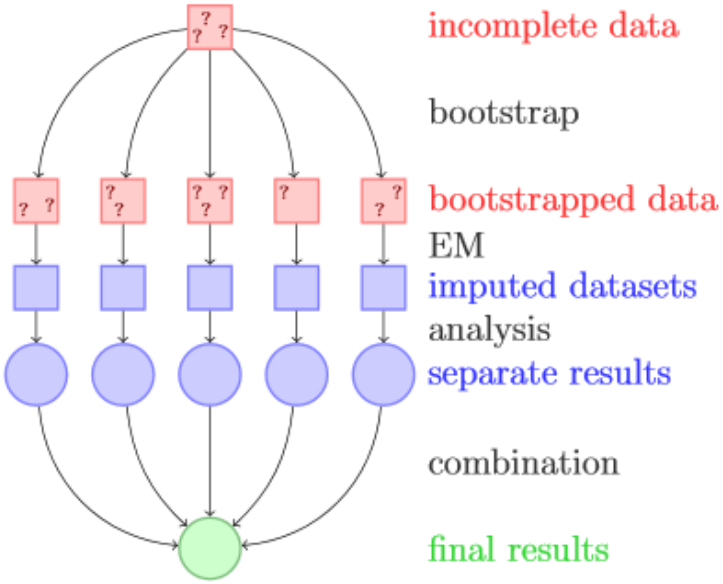


Figure 1. Amelia II schematic approach to multiple imputation with the expectation-maximization with bootstrapping (EMB) algorithm. Figure taken from the Amelia II manual (Honaker, King & Blackwell, 2015)

The expectation-maximization (EM) algorithm (Dempster, Laird & Rubin, 1997) is a general-purpose computational approach to finding the mode of the posterior. In their original form the EM estimates cannot be used to create MI, as the estimates do not reflect the fact that they have been estimated from a finite sample. In order to solve this, Amelia II first takes *m* bootstrap samples, and applies the EM-algorithm to each of these bootstrap samples (Figure 1). The *m* estimates of means and variances will now be different. The first set of estimates is used to draw the first set of imputed values by a form of regression analysis, the second is used to calculate the second set of imputed values, and so on. The bootstrapping creates samples based on the distribution, and the EM-algorithm fits those samples on the original data. The decision for which value will be put in place of the missing value is made by regression. The imputed value is calculated as the predicted value plus a random draw from the residual distribution (King, Tomz and Wittenberg, 2000).

MI involves imputing *m* values for each missing cell in the data matrix and creating *m* 'completed' data sets. Across these completed data sets, the observed values are the same, but the missing values are filled in with a distribution of imputations that reflect the uncertainty about missing data. After imputation with Amelia II's EMB algorithm, all statistical methods can be applied that would have been used if there had been no missing values to each of the *m* data sets, and the results from *m* analyses can be combined in interpretation. For the purposes of this study, the program default of m=5 has been used, i.e. five new imputed datasets were created for each analysis.

2.3.2. Imputation assumptions

The imputation model in Amelia II assumes that the data are missing at random (MAR) and not missing completely at random (MCAR). This assumption means that the pattern of missingness only depends on the observed data, not the unobserved data. For such data multiple imputation is deemed suitable, once the reasons for missingness are included into the EMB model.

The second MI assumption is that the complete data (that is, both observed and unobserved) are multivariate normal, so all information about the relations in the data can be summarized by just mean and covariances. When data are incomplete, Amelia II uses the EM algorithm to find corrected

estimates of the means and covariances (Little & Rubin, 2002). However, there is also evidence that this model works as well even with categorical or mixed data (Schafer & Olsen, 1998).

2.3.3. Imputation methodology

For the MI as a prior all unobserved values on cognitive tests were set as positive (minimum value of 0) and the highest logical value on that variable for all phases and all participants was set as a maximum.

All imputations were performed on continuous data; only after imputation the variables were transformed into binary format ('0' not impaired and '1' impaired) based on the same thresholds as had been done in Chapter Five (Table 5.2).

2.4. Dimensionality analysis

In this supplementary insert the dimensionality analyses were carried out following the same methodology as presented in the methods section of Chapter Five.

This insert starts with an attempt to investigate the domain-specificity of cognitive disturbances. The second but linked analysis revolved around identification of potential subgroups (clusters) of individuals (within the study population) who display distinct patterns of cognitive impairments and/or cognitive change. The dimensionality of the phase I and phase II datasets were analysed separately, then later the differences between phase I and phase II dimensionality were analysed as a third (longitudinal change) dataset.

All analyses with MI have been be performed five times as five MI datasets, as per recommendations of the methodology developers (Honaker et al., 2015).

3. Results

3.1. Participant data completeness

Of the 82 patients with both phase I and phase II data, 21 had missing observations on at least one of the sixteen cognitive variables at either phase I or at phase II (Table 1). In Chapter Five analyses these individuals were removed by performing list-wise deletion, however, in this insert, these individuals were included. This was done after attributing the causes of missingness, therefore properly managing the missing data.

Table 1. Cognitive data completeness at phases I and II

	Missing observations	
	Phase I	**Phase II**
SRTL	1/82	1/82
SRTC	1/82	1/82
SRTD	1/82	1/82
SPART	1/82	4/82
SPARTD	1/82	4/82
SDMT	3/82	1/82
PASAT	**9/82**	**16/82**
WLGT	0/82	0/82
VOSP2	2/82	0/82
VOSP3	2/82	1/82
VOSP6	2/82	3/82
VOSP8	2/82	1/82
BORB2	3/82	1/82
BORB5	3/82	3/82
TROG	3/82	1/82
BC.Index	**12/82**	2/82

Table 1 shows that the PASAT and BC.Index (at phase I) variables had the highest numbers of missing observations. Missing observations on other variables were minimal.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

On both phases there had been high numbers of missing observations on the PASAT test (11% phase I and 19.5% phase II). This indicated that the instructions of the PASAT test had been too difficult for many patients to understand, and because they didn't know how to perform the test, these pwMS ended up not doing it at all.

The high number of missing observations on the BC test at phase I (14.6%) were because this test was added to the study battery later on, after the study had already started. For the most part, the missing observations indicated the individuals that were the first ones to join the phase I study.

The individual patterns of missing data for phase I can be seen in Figure 2; for phase II in Figure 3; and the comparison of missingness between the two phases can be seen in Figure 4.

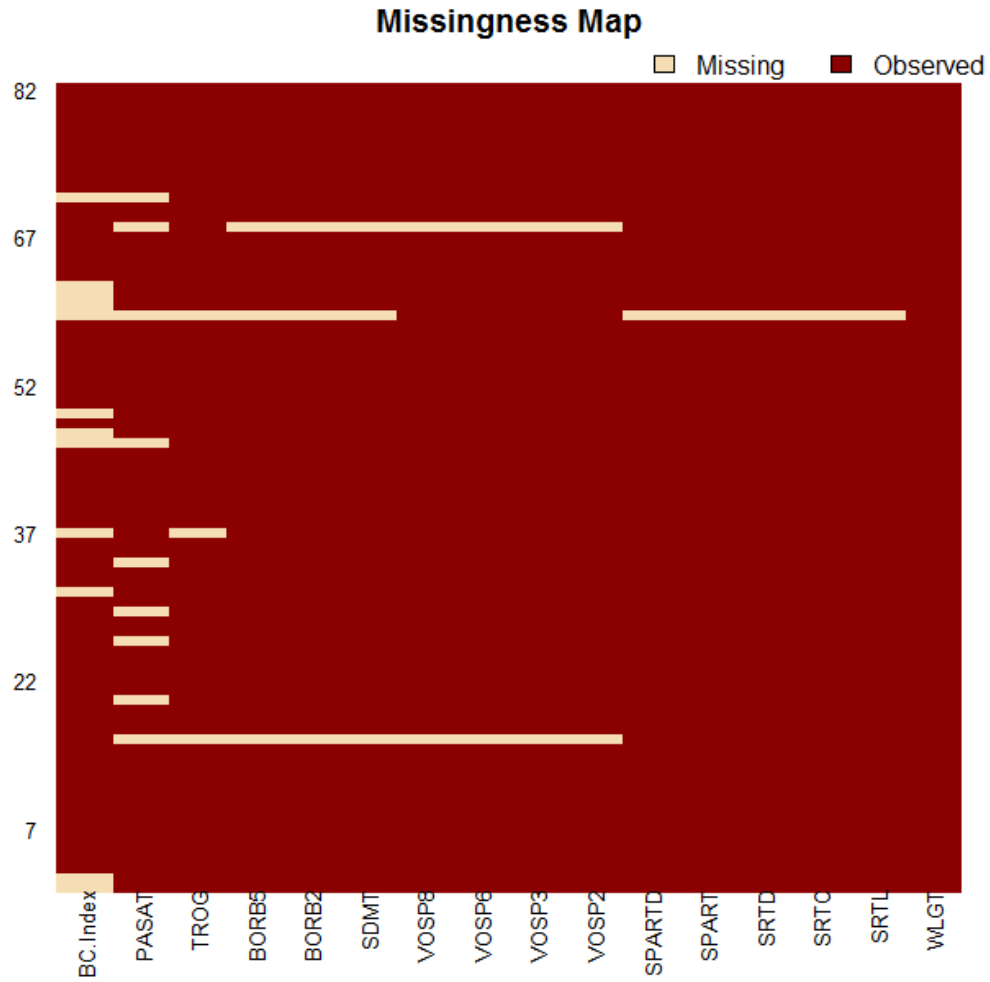


Figure 2. Missingness map of the phase I cognitive dataset (n=82).

Note. Missing values are in tan and observed values are in red. The columns represent the variables and the rows represent individual patients. The variables are arranged in the order from most to least missing values from left to right. The rows are in random order.

Figure 2 shows that at phase I the pattern of missing observations was not random, as there were three individuals with high numbers of missing observations on many of the variables. All other patients with missing observations had data missing on the BC.Index and PASAT variables.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

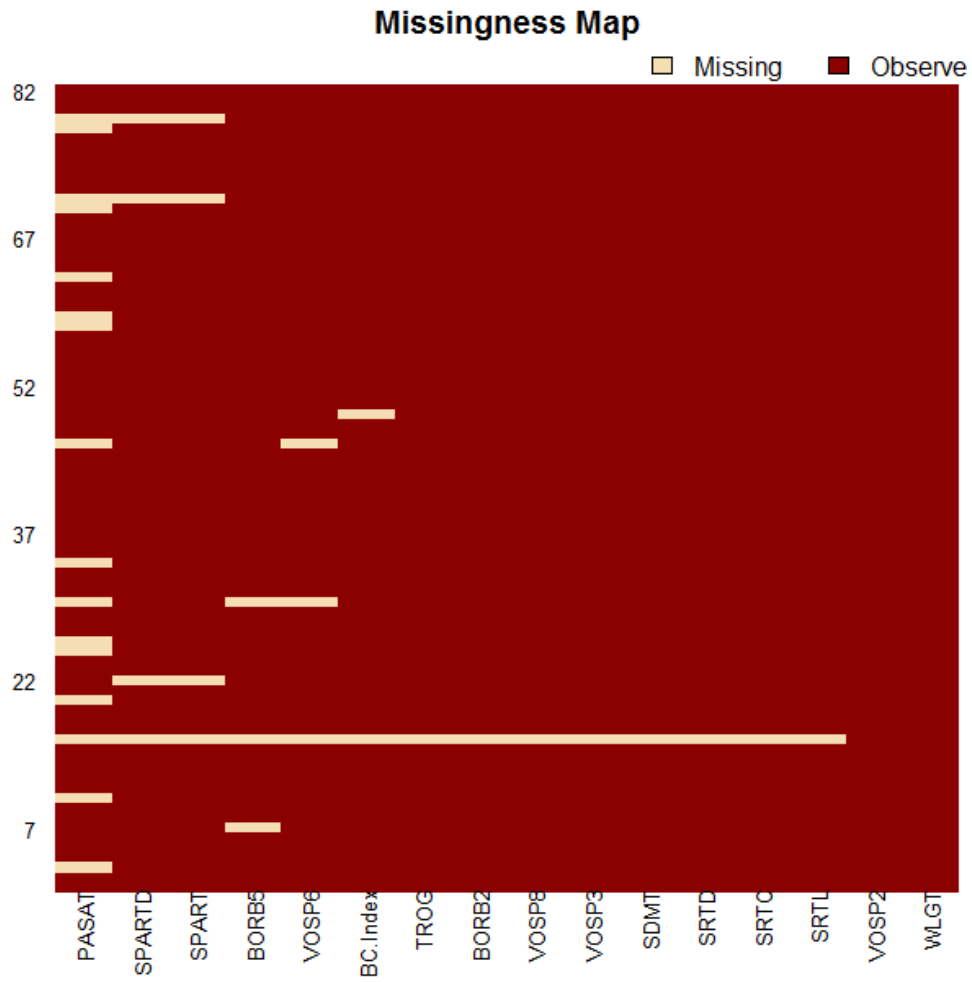


Figure 3. Missingness map of the phase II cognitive dataset (n=82).

Note. Missing values are in tan and observed values are in red. The columns represent the variables and the rows represent individual patients. The variables are arranged in the order from most to least missing values from left to right. The rows are in random order.

Figure 3 shows that at phase II there was one individual with most of the data missing, a few individuals with some data missing on the visuoperceptual tests, and the majority of missing observations were on the PASAT tests.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

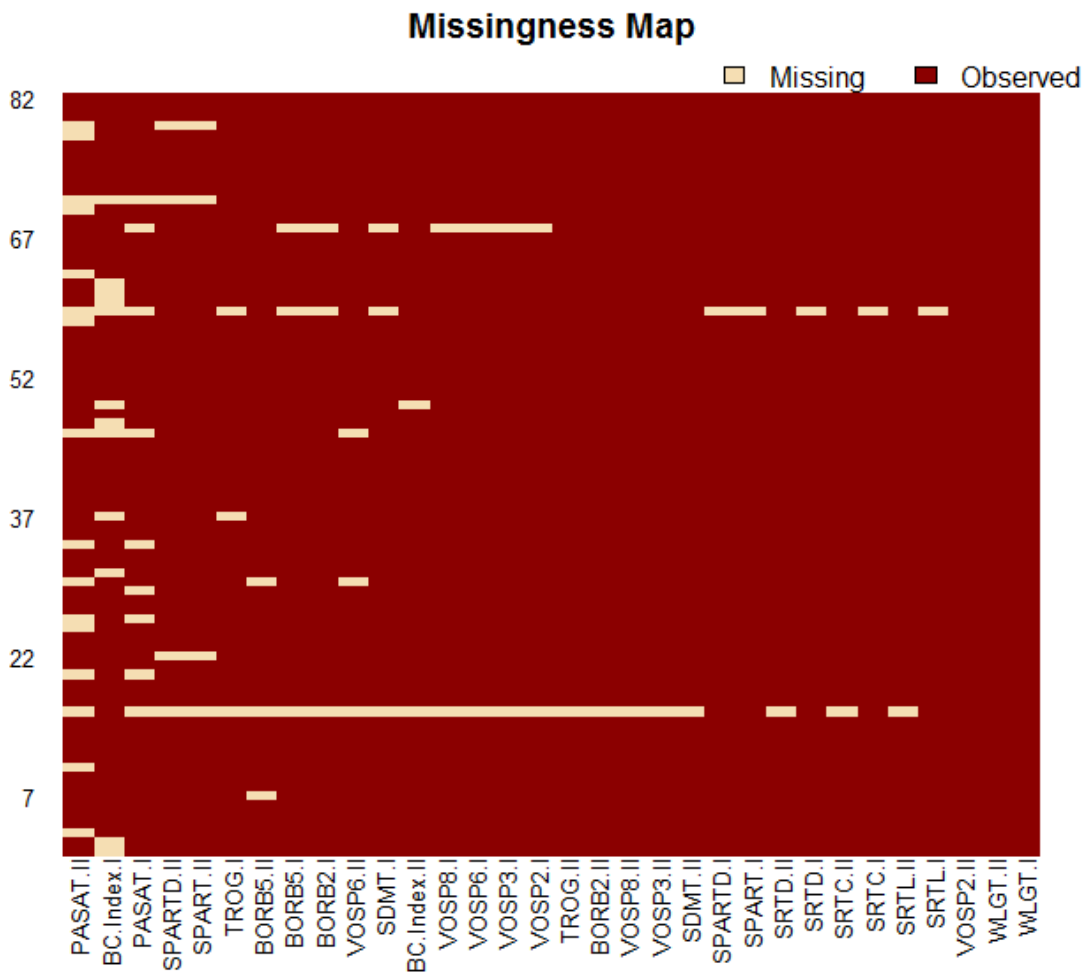


Figure 4. Missingness map of the phases I and II cognitive datasets (n=82).

Note. Missing values are in tan and observed values are in red. The columns represent the variables and the rows represent individual patients. The variables are arranged in the order from most to least missing values from left to right. The rows are in random order. The number at the end of the name each variable on the x-axis indicates at which phase that data was collected (e.g. PASAT.I was collected at phase I, and PASAT.II was collected at phase II).

Figure 4 shows that just a few patients had missing data on the same variables on both phases (e.g. PASAT I and PASAT II; BORB5 I and BORB5 II). There was one individual who was unable to perform many of the items on both phases; one individual who had visuoperceptual data missing at phase I, but not phase II; and one individual who had memory data missing at phase I, but not phase II.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

At phase I it seemed that the data was not MCAR – several certain individuals had more missing data than others (Figure 2). Specifically, two individuals had completed all but the visuoperceptual tests, and one individual had missing data on all but the VOSP items. Besides these two cases, the remaining individuals had data missing only on the BC test and PASAT.

At phase II again the data was not MCAR (Figure 3), as it can be seen that most individuals had data missing on the PASAT test. A few patients had data missing on the visuoperceptual tests, and only one individual had a large number of observations missing, having completed only the VOSP2 and WLGT tests.

As it can be seen from Figure 4, just a few individuals had missing data on the same variables at both phases. One patient had many missing observations due to physical limitations, however, for that specific participant more observations were missing at follow-up than at baseline. This patient's MS had severely progressed, limiting her to a totally helpless bed patient who was unable to communicate effectively (phase I EDSS=8; phase II EDSS=9.5).

The PASAT was the only test that had large numbers of unobserved values on both phases. This indicated that the individuals with unobserved PASAT values were unable to understand it's instructions during both assessments, and that the missing values were specific to the test, rather than to the administration protocol differences between the phases I and II.

As has been shown in Table 1 and Figures 2-4, in this study the missing observations can be considered to be missing at random (MAR), and not missing completely at random (MCAR), as the sources of all missingness can be explained in this study. The main sources of missingness were inability to move the hands for missing SPART observations; complicated instructions and cognitive difficulties in understanding instructions for PASAT; visual problems for missing values on some of the visuoperceptual items; the order of recruitment for BC Test missing data at phase I; etc. Therefore the study dataset meets the MI assumptions and thus the missing values can be imputed with the MI method.

3.2. Individuals with complete and missing data

Using the MI method the missing values were imputed for 21 individuals.

In order to investigate whether there had been any systematic differences between individuals with complete and missing data, their clinical and demographic characteristics were examined.

Table 2. Demographic and clinical characteristics of the individuals with complete and missing data

	Complete data (n=61)	**Missing observations (n=21)**	**Difference**
Age	50.23 (9.18)	53.30 (9.69)	U = 750, p = 0.125
Gender (%)	53% Female, 47% male	55% female, 45% male	X^2 = 0.039, p = 0.843
Years of education	13.05 (3.15)	12.11 (2.85)	U = 455, p = 0.252
Dementia (ACE-R < 82)	89% not impaired, 11% impaired	85% not impaired, 15% impaired	X^2 = 0.173, p = 0.678
BDI-II	15.25 (9.22)	15.32 (10.66)	U = 577.50, p = 0.982
MSIS-29	82.97 (25.99)	85.16 (28.71)	U = 602, p = 0.799
MS course (%)	26% PPMS, 30% SPMS, 44% RRMS	14% PPMS, 48% SPMS, 38% RRMS	X^2 = 2.602, p = 0.272
Disease duration	**12.58 (6.81)**	**17.85 (9.92)**	**U = 793, p = 0.032***
EDSS	5.16 (1.83)	5.80 (2.33)	U = 703.50, p = 0.282

Note. Data for comparison of demographic and clinical characterisics was collected at phase II. Comparisons were performed using the Mann-Whitney U-test for interval and ordinal data, and Pearson's Chi Square test for frequency data

Table 2 shows that in general, there had been no differences between the patients with complete observations and those with some missing data, however, the missing observations were associated with longer MS duration

Abbreviations: ACE-R – Addenbrooke's Cognitive Examination – Revised, BDI-II – Beck's Depression Inventory 2nd Ed., MSIS-29 – 29-item Multiple Sclerosis Impact Scale, EDSS – Expanded Disability Status Scale

As it can be seen from Table 2, the only differences between the pwMS with complete data (n=61) and missing data (n=21) were in the disease duration, therefore it appears that the participants who had missing observations have had MS for longer. However, such difference was not unexpected, as the reason why the data was not collected had in part to do with the patients' more severe MS symptoms (such as visual problems or tremor); and the longer someone has had MS, the more likely they were to develop more symptoms that may have interfered with cognitive testing.

Besides the duration of MS, there were no other differences between the patients included in the Chapter Five analyses, and those who were not.

3.3. Imputation

This study has met all of the assumptions necessary for MI, therefore it was proceeded with imputing the missing values.

In order to improve the accuracy of MI, it was considered to run the imputation procedures separately for the phase I and phase II data sets. First, in order to better predict performance of severely disabled individuals, it was decided to include the EDSS scores among the information variables into both data sets. Moreover, since the BC.Index missing values occurred at baseline, they could be partially predicted by follow-up data values. Therefore for phase I, in order to have better accuracy for BC.Index data imputation, in addition to EDSS scores, the BC.Index scores from phase II were also included among the information variables.

However, since the PASAT and the BC.Index items had the most missing observations, in further dimensionality analyses it was important to keep in mind that these items had the most data 'created' therefore their results should be interpreted with caution.

3.4. The dimensionality of cognitive impairments

The result analysis of this chapter begins with an attempt to investigate the pattern in which the cognitive functions are affected by MS. First the dimensionality of cognitive impairments of phase I and phase II datasets was analysed separately, then later the dimensionality of emerging and resolving deficits as a third longitudinal dataset was studied.

3.4.1. Domain-specificity of impairments at phase I

The first step was to investigate the pattern of domain deficits at phase I. This analysis of dimensionality allowed to explore two elements – the cumulative burden of impairments (i.e. were they predominantly single-domain, or multi-domain), and the pattern of impairments (i.e. was there obvious over-representation of certain combinations of cognitive domain failure).

Table 3. The pattern of domain deficits for pwMS at phase I. Complete case analysis (n=61) and multiple imputation (m=5; n=82)

Cognitive domain	Number of items in each domain	Minimum number of tests to fail to be considered impaired on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

0 Domain unaffected. Represents performance below cut-off on one or less tests in that domain.

1 Domain impaired. Represents performance below cut-off on two or more tests in the domain.

						Number of pwMS (of 82)					
Total domain impairments	**Verbal memory**	**Visuospatial memory**	**Processing speed**	**Visuo-perceptual**	**Language**	**Complete case analysis (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
No impairments	0	0	0	0	0	**33 (54%)**	**39 (48%)**	**39 (48%)**	**39 (48%)**	**39 (48%)**	**39 (48%)**
Impairment on only one domain	1	0	0	0	0	8 (13%)	14 (17%)	13 (16%)	14 (17%)	15 (18%)	13 (16%)
	0	1	0	0	0	2 (3%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)
	0	0	1	0	0	1 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)
	0	0	0	1	0	4 (7%)	8 (10%)	7(9%)	7 (9%)	8 (10%)	8 (10%)
	0	0	0	0	1						
Impairment on two domains	1	1	0	0	0						
	1	0	1	0	0	4 (7%)	5 (6%)	5 (6%)	5 (6%)	5 (6%)	5 (6%)
	1	0	0	1	0	4 (7%)	6 (7%)	6 (7%)	7 (8%)	5 (6%)	6 (7%)
	1	0	0	0	1	1 (2%)	1 (1%)	2 (2%)	1 (1%)	1 (1%)	2 (2%)
	0	1	0	0	1						
	0	0	1	1	0	1 (2%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	0	0	1	0	1						
	0	0	0	1	1	1 (2%)	1 (1%)	2 (2%)	1 (1%)	1 (1%)	1 (1%)
Impairment on three domains	1	1	1	0	0						
	1	1	0	1	0						
	1	1	0	0	1						
	1	0	1	1	0						
	1	0	1	0	1	1 (2%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
Impairment of four domains	1	1	1	1	0	1 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)
	1	1	1	0	1						
Global impairment	1	1	1	1	1						

Table 3 shows that at phase I around half of the patients had no cognitive domains failed, a quarter had one domain failed, approximately one sixth had two domains failed, and a very small number of patients had multidimensional impairments.

As it can be seen from Table 3, at phase I the profiles of patients with complete cases were non-different from those with imputed data. However, the sole minor difference can be observed in the prevalence of patients without any impairments – there were slightly less individuals with no impairments among the patients with missing data, although this difference was quite small (approx. 6%).

Again, as had been shown in Chapter Five, at phase I the patients tend to fail all of the cognitive domains, although the impairments on verbal memory items were the most common. Besides that there seemed to be no over-representation of a particular combination, suggesting that cognitive deficits at phase I could indeed be considered multidimensional. At phase I most patients failed only one or two cognitive domains, and only a few patients could be considered more globally impaired having failed three or four cognitive domains.

To sum up, at phase I the findings from the imputed samples were non-different from those presented in the complete case analysis in Chapter Five.

3.4.2. Domain-specificity of impairments at phase II

The next step was to investigate whether the MI has caused any differences in the dimensionality of cognitive impairments at phase II. Again, at phase II the MI data had the same dimensionality as the complete case analysis data, except that after MI more patients have had severe global impairments (Table 4). These were the only cases that had been missed with the complete case analysis. Besides that, no differences were found between the complete case and MI data dimensionality analyses at phase II.

In phase II patients had developed more multi-dimensional deficits, as more patients have failed three or more cognitive domains than in phase I.

At phase II the multidimensional impairments had similar patterns – these patients were more likely to have failed verbal memory, processing speed and visuoperceptual groups of tests. The patients who had failed four cognitive domains had also had the same pattern with an additional impairment of visuospatial memory.

Table 4. The pattern of domain deficits for pwMS at phase II. Complete case analysis (n=61) and multiple imputation (m=5; n=82)

Cognitive domain	Number of items in each domain	Minimum number of tests to fail to be considered impaired on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

0 Domain unaffected. Represents performance below cut-off on one or less tests in that domain.

1 Domain impaired. Represents performance below cut-off on two or more tests in the domain.

						Number of pwMS (of 82)					
Total domain impairments	**Verbal memory**	**Visuospatial memory**	**Processing speed**	**Visuo-perceptual**	**Language**	**Complete case analysis (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
No impairments	0	0	0	0	0	27 (44%)	31 (38%)	31 (38%)	31 (38%)	32 (39%)	31 (38%)
Impairment on only one domain	1	0	0	0	0	3 (5%)	5 (6%)	5 (6%)	5 (6%)	5 (6%)	5 (6%)
	0	1	0	0	0	3 (5%)	3 (4%)	3 (4%)	3 (4%)	3 (4%)	3 (4%)
	0	0	1	0	0	7 (11%)	10 (12%)	10 (12%)	10 (12%)	9 (11%)	10 (12%)
	0	0	0	1	0	2 (3%)	5 (6%)	5 (6%)	5 (6%)	4 (5%)	4 (5%)
	0	0	0	0	1						
Impairment on two domains	1	1	0	0	0		1 (1%)				
	1	0	1	0	0	3 (5%)	3 (4%)	5 (6%)	5 (6%)	5 (6%)	4 (5%)
	1	0	0	1	0	2 (3%)	3 (4%)	3 (4%)	3 (4%)	3 (4%)	3 (4%)
	1	0	0	0	1						
	0	1	1	0	0				1 (1%)		1 (1%)
	0	1	0	1	0	1 (2%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	0	1	0	0	1						
	0	0	1	1	0	3 (5%)	5 (6%)	5 (6%)	6 (7%)	7 (8%)	7 (8%)
	0	0	1	0	1						
	0	0	0	1	1						
Impairment on three domains	1	1	1	0	0	1 (2%)	2 (2%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	1	1	0	1	0						
	1	1	0	0	1						
	1	0	1	1	0	8 (13%)	9 (11%)	8 (10%)	8 (10%)	8 (10%)	9 (11%)
	1	0	1	0	1						
	1	0	0	1	1						
	0	1	1	1	0			2 (2%)		1 (1%)	
Impairment of four domains	1	1	1	1	0	**1 (2%)**	**4 (5%)**	**3 (4%)**	**3 (4%)**	**3 (4%)**	**3 (4%)**
	0	1	1	1	1						
Global impairment	1	1	1	1	1						

Table 4 shows that at phase II less than half of pwMS were considered to be cognitively spared; around 15% had one or two domains impaired, with overall higher prevalence of multidimensional impairments than at phase I.

3.4.3. Dimensionality of temporal change in cognitive impairments

In the previous analyses the patterns were shown of how cognitive impairments group in pwMS at two separate phases of the assessment. However, what was left unexplained was how stable those impairments were, and what was the pattern of their longitudinal development.

The dimensionality of change was analysed in two ways and separate steps were taken to investigate the impairments that emerge, and the impairments that resolve. This was done in order to explore whether any patterns of domain-specificity of evolving or resolving impairments could be observed.

Several measures were undertaken in order to avoid the effect of potential artefacts in the dimensionality analysis that could potentially limit the generalizability of the findings. I identified a possibility that there could also be individuals that show patterns of both emerging and resolving impairments in the same domain, but on different tests. Because of this reason it was chosen to define longitudinal change on a cognitive domain when there were at least two observations on two separate tests in the same direction, following the same reasoning as per definition of impairment in Table 3 and Table 4. This adjustment made the analyses of longitudinal change more conservative, especially on the domains that have fewer tests in them and was shown to be successful in Chapter Five in reducing the risk of Type I error.

I. Dimensionality of emerging impairments

The first part of longitudinal analyses of changes in dimensionality of cognitive impairment revolved around identification of evolving impairments. These were the instances where a patient has performed at norm at phase I, but below the cut-off at phase II (Table 5). In majority, the MI datasets showed similar patterns as the complete case analysis. Around half of patients did not develop any new domain impairments, at least none that could get picked up by this conservative definition.

However, there have also been differences between the complete case and MI analyses. In the complete case analysis almost all of the newly developed domain impairments were uni-dimensional, with emerging impairments in processing speed and visuoperception most prominent.

In the analysis of the dimensionality of emerging deficits in the MI datasets, there have been a few patients who have suddenly developed more global impairments with emerging deficits in two or three domains.

Table 5. Dimensionality of developing impairments. The pattern of acquired deficits for pwMS in each domain between phase I and phase II is presented. Complete case analysis (n=61) and multiple imputation (m=5; n=82)

Cognitive domain	Number of items in each domain	Minimum number of items to develop new impairments on to be considered having emerging impairments on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

0 Domain unaffected. Represents performance below cut-off on one or less tests in a domain.

1 Domain impaired. Represents performance below cut-off on two or more tests in the domain.

						Number of pwMS (of 82)					
Total new domain impairments	**Verb. mem**	**Vis. mem**	**Proc. speed**	**Vis.perc**	**Language**	**Complete case analysis (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
No new impairment	0	0	0	0	0	39 (64%)	42 (51%)	41 (50%)	41 (50%)	42 (51%)	41 (50%)
New impairment on only one domain	1	0	0	0	0	2 (3%)	4 (5%)	4 (5%)	4 (5%)	4 (5%)	4 (5%)
	0	1	0	0	0	2 (3%)	3 (4%)	3 (4%)	3 (4%)	3 (4%)	3 (4%)
	0	0	1	0	0	6 (10%)	10 (12%)	13 (16%)	13 (16%)	12 (15%)	12 (15%)
	0	0	0	1	0	7 (11%)	9 (11%)	10 (12%)	10 (12%)	9 (11%)	9 (11%)
	0	0	0	0	1						
New impairment on two domains	1	1	0	0	0		3 (4%)				
	1	0	1	0	0	1 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)	2 (2%)
	1	0	0	1	0	1 (2%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	1	0	0	0	1						
	0	1	1	0	0			2 (2%)	2 (2%)	2 (2%)	2 (2%)
	0	1	0	1	0						
	0	1	0	0	1						
	0	0	1	1	0	3 (5%)	5 (6%)	3 (4%)	3 (4%)	4 (5%)	5 (6%)
	0	0	1	0	1						
	0	0	0	1	1						
New impairment on three domains	1	1	1	0	0		1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	1	1	0	1	0						
	1	1	0	0	1						
	1	0	1	1	0		1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	1	0	1	0	1						
	1	0	0	1	1						
	0	1	1	1	0		1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	0	1	1	0	1						
	0	1	0	1	1						
New impairment on four domains	1	1	1	1	0						
	1	1	1	0	1						
	1	1	0	1	1						
	1	0	1	1	1						
	0	1	1	1	1						

Table 6. Dimensionality of resolving impairments. The pattern of resolving deficits for pwMS in each domain between phase I and phase II is presented. Complete case analysis (n=61) and multiple imputation (m=5; n=82)

Domain	Number of tests in each domain	Minimum number of tests to improve on to be considered having resolving impairments on each domain
Verbal memory	3 (SRTL, SRTC, SRTD)	2 / 3
Visuospatial memory	2 (SPART, SPARTD)	2 / 2
Processing speed	3 (SDMT, PASAT, WLGT)	2 / 3
Visuoperceptual	6 (VOSP2, VOSP3, VOSP6, VOSP8, BORB2, BORB5)	2 / 6
Language	2 (TROG, BC.Index)	2 / 2

1 Impairments resolve. I.e. impaired at phase I and non-impaired at phase II

0 New impairments that emerge or remain stable.

						Number of pwMS (of 82)					
Total domain recovery	**Verbal memory**	**Visuospatial memory**	**Processing speed**	**Visuo-perceptual**	**Language**	**Complete case analysis (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
No recovery	0	0	0	0	0	51 (84%)	62 (76%)	59 (72%)	59 (72%)	62 (76%)	59 (72%)
Recovery on only one domain	1	0	0	0	0	4 (7%)	10 (12%)	11 (13%)	11 (13%)	10 (12%)	10 (12%)
	0	1	0	0	0	1 (2%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	0	0	1	0	0						
	0	0	0	1	0	2 (3%)	5 (6%)	5 (6%)	6 (7%)	4 (5%)	6 (7%)
	0	0	0	0	1	1 (2%)	1 (1%)	3 (4%)	1 (1%)	1 (1%)	2 (2%)
Recovery on two domains	1	1	0	0	0				1 (1%)	1 (1%)	1 (1%)
	1	0	1	0	0						
	1	0	0	1	0	2 (3%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	1	0	0	0	1		1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
	0	1	1	0	0						
	0	1	0	1	0						
	0	1	0	0	1						
	0	0	1	1	0						
	0	0	1	0	1						
	0	0	0	1	1		1 (1%)	1 (1%)	1 (1%)	1 (1%)	1 (1%)
Recovery on three domains	1	1	1	0	0						
	1	1	0	1	0						
	1	1	0	0	1						
	1	0	1	1	0						
	1	0	1	0	1						
	1	0	0	1	1						
	0	1	1	1	0						
	0	1	1	0	1						
	0	1	0	1	1						
Recovery on four domains	1	1	1	1	0						
	1	1	1	0	1						
	1	1	0	1	1						
	1	0	1	1	1						
	0	1	1	1	1						

Sudden global impairment	1	1	1	1	1					

Table 5 shows that around half of the patients don't develop new impairments in the time interval between phase I and phase II. About a third of patients have developed new impairments on one domain, 10% on two domains, and it was very unlikely to suddenly develop multidimensional impairments in 3 year follow-up.
Table 6 shows that for the majority of patients the impairments exhibited at phase I didn't resolve. Around 20% of patients had impairments resolved on one domain, with most improvement seen on verbal memory. Very small numbers of patients showed recovery on two domains indicating that a multidimensional recovery is highly unlikely.

II. Dimensionality of resolving impairments

The vast majority of patients did not improve on any of the cognitive domains, and if they did, they mostly improved on only one domain (Table 6). This was consistent to both complete cases and MI datasets. Therefore it could be considered, that in general, the changes between phases I and II were not random, instead, a trend for acquiring new impairments was observed, with most of new problems emerging on visuoperceptual and information processing speed tests.

3.4.4. Dimensionality of cognitive impairments. Summary

From analysing the domain-specificity of cognitive impairments it was found that all cognitive domains can be affected in MS, although not at the same frequencies. The majority of individuals from the phase I MS sample had deficits only on one cognitive domain, however, with when re-assessed three years later they became impaired on more cognitive domains, with emerging deficits on visuoperceptual and information processing speed tests being most prominent.

From the comparison of complete cases and MI datasets it can be summarised that in general both methods of data analysis have provided similar results, however, the MI datasets included slightly more patients with more global impairments and smaller prevalence of patients without any impairments; which was expected as physical disability caused by more severe disease progression had interfered with collection of cognitive data.

In the analysis of the dimensionality of emerging deficits in the MI datasets, there have been a few patients who have suddenly developed more global impairments with emerging deficits in two or

three domains. The progression profiles exhibited by these three individuals represented a more severe deterioration and had not been present in the initial complete case analyses in Chapter Five.

3.5. The dimensionality of the phase I study population

After completing the analysis of the dimensionality of cognitive impairments, the next step was to investigate the dimensionality of the study cohort in order to address the emerging patterns of grouping the pwMS based on their performance. With this analysis we aimed to understand whether certain groups of patients that share similar patterns of cognitive performance could be identified.

Again, all cluster analyses were performed in total of six times – first as the complete cases analysis, as had been done in Chapter Five, then five times with all MI datasets.

As had been done in Chapter Five, the subject-cluster analyses were first performed on the phase I datasets, then on the phase II datasets, and then finally on the datasets of longitudinal change.

3.5.1. Subject-clusters in the phase I dataset

First we examined the phase I dataset to find the optimum number of clusters of patients based on their cognitive performance. We separately sought to identify the most suitable number of clusters in each dataset – the complete cases dataset and then separately in the five imputed datasets (Table 7).

Table 7. Determination of the optimal number of clusters in the hierarchical cluster analysis of patient data at phase I (n=82)

	Caliński-Harabasz pseudo-F					
Number of Clusters	**Complete cases (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
2	3.39	6.66	5.53	6.06	4.78	6.14
3	3.59	5.19	4.57	5.14	4.20	4.90
4	5.28	3.98	3.61	4.23	3.99	4.44
5	4.6	3.60	4.01	3.74	3.52	3.86
6	4.22	**7.17**	3.65	3.44	3.11	3.42
7	4.72	7.08	3.29	3.24	3.02	5.71
8	6.22	6.60	3.60	3.38	3.64	5.48
9	6.29	6.18	3.44	3.51	5.97	5.66
10	**6.33**	6.51	5.60	3.48	6.27	5.78
11	6.15	6.92	5.80	6.07	6.45	6.52
12	5.9	6.70	5.78	5.72	6.67	6.69
13	6.1	6.71	**6.18**	**6.09**	6.62	6.40
14	5.65	6.90	6.04	5.97	6.76	**6.74**
15	5.61	6.72	5.84	5.74	**6.87**	6.52

Table 7 shows that there were no clear optimum number of clusters in the phase I dataset. The complete case analysis and the five MI datasets suggested that the number of clusters ranged in-between six and fifteen.

According to the Caliński-Harabasz cluster stopping criterion, there is no one clear optimum number of clusters at phase I data. Depending on the dataset (complete cases or the five MI), the number of subject-clusters for the phase I sample ranged between six and fifteen. Moreover, as can be seen from the Table 8 and the dendrograms (Figure 5), in each clustering solution the clusters were not equal in their sizes. In each solution there tended to be one largest cluster that grouped around half of patients, leaving small numbers of patients in the remaining clusters. This pattern with half of patients in cluster one, and with another half distributed in the remaining (five to fourteen) clusters, was present in all six datasets.

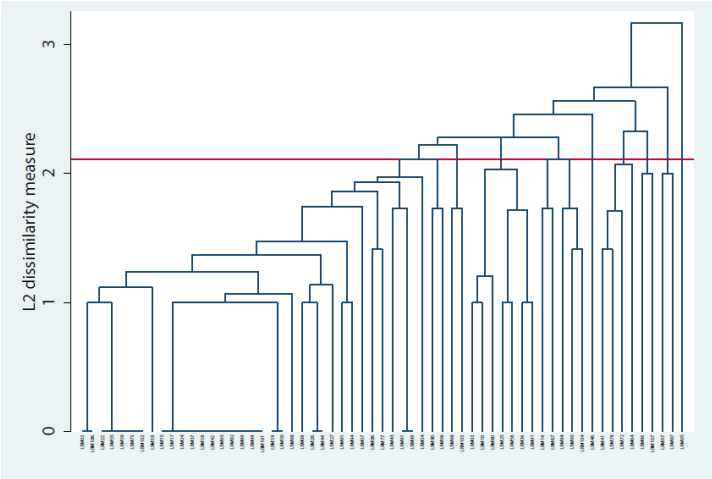
Table 8. Number of individuals in each subject-cluster solution at phase I (n=82)

Cluster	Complete cases (n=61)	Imputation 1	Imputation 2	Imputation 3	Imputation 4	Imputation 5
1	35 (57%)	48 (59%)	41 (50%)	41 (50%)	38 (46%)	40 (49%)
2	2 (3%)	27 (33%)	4 (5%)	16 (20%)	3 (4%)	4 (5%)
3	2 (3%)	2 (2%)	13 (16%)	8 (10%)	4 (5%)	5 (6%)
4	7 (12%)	2 (2%)	3 (4%)	2 (2%)	6 (7%)	1 (1%)
5	5 (8%)	2 (2%)	6 (7%)	3 (4%)	3 (4%)	3 (4%)
6	1 (2%)	1 (1%)	4 (5%)	3 (4%)	5 (6%)	2 (2%)
7	4 (7%)		2 (2%)	1 (1%)	6 (7%)	8 (10%)
8	2 (3%)		1 (1%)	1 (1%)	4 (5%)	7 (9%)
9	2 (3%)		3 (4%)	1 (1%)	4 (5%)	4 (5%)
10	1 (2%)		2 (2%)	2 (2%)	2 (2%)	1 (1%)
11			1 (1%)	1 (1%)	1 (1%)	2 (2%)
12			1(1%)	1 (1%)	2 (2%)	2 (2%)
13			1 (1%)	2 (2%)	2 (2%)	1 (1%)
14					1 (1%)	2 (2%)
15					1 (1%)	

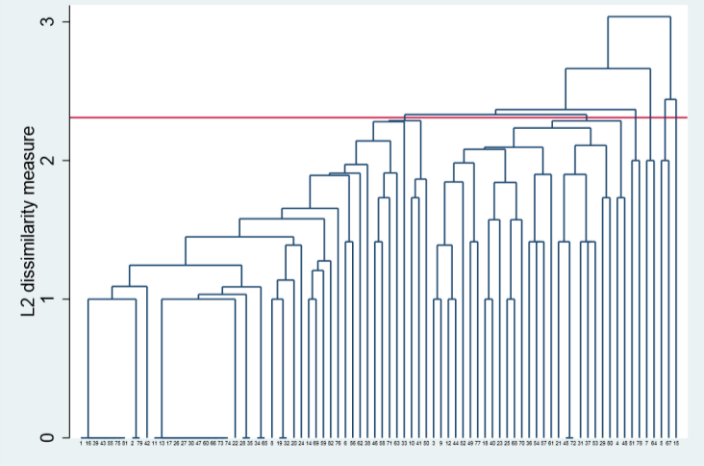
Table 8 shows that at phase I around half of patients shared similar patterns of cognitive performance, and the other half of patients were split between the remaining five to fourteen cluster.

Figure 5. Dendrogram of hierarchical cluster analysis using average linkage model of phase I patient data. The horizontal line is shown to represent where the optimal cluster cut falls in each analysis

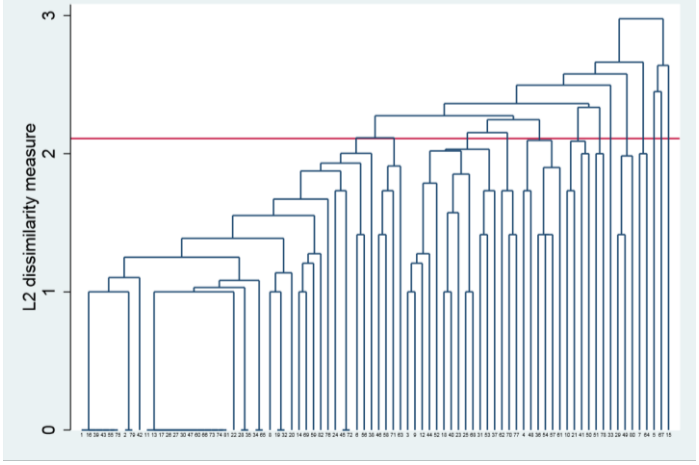
a) Phase I, Complete cases (n=61). Ten-cluster solution



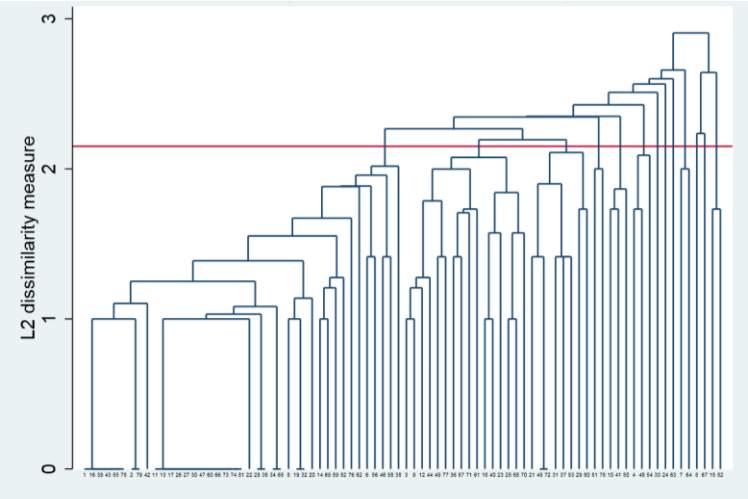
b) Phase I, Imputation 1 (n=82). Six-cluster solution



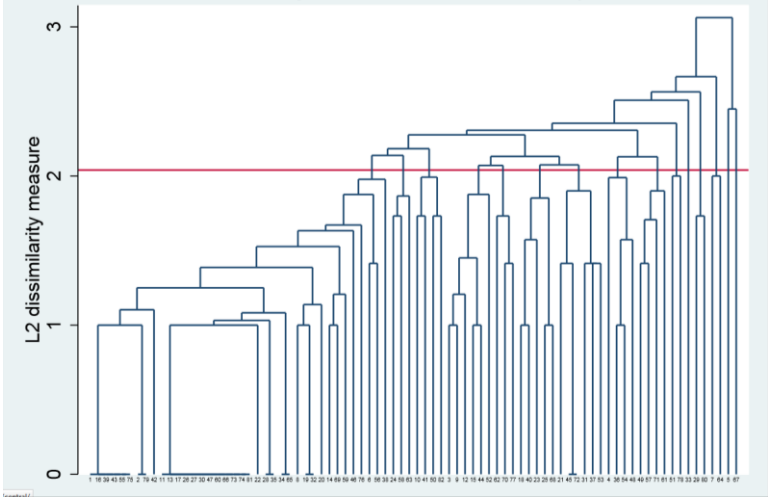
c) Phase I, Imputation 2 (n=82). Thirteen-cluster solution



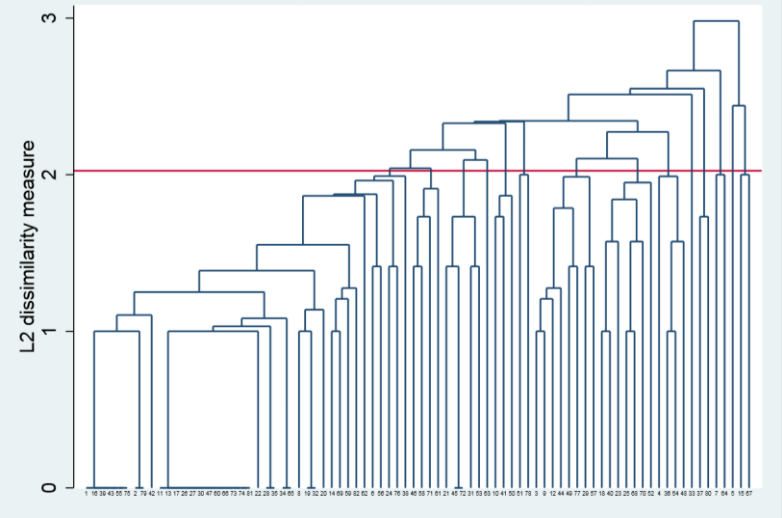
d) Phase I, Imputation 3 (n=82). Thirteen-cluster solution



e) Phase I, Imputation 4 (n=82). Fifteen-cluster solution



f) Phase I, Imputation 5 (n=82). Fourteen-cluster solution



After splitting the phase I patients into groups, the next step was to investigate the cognitive profiles of the individuals in each cluster. As it can be seen from Table 9, all of the phase I datasets (complete cases and all imputations) have yielded similar results. At phase I half of the patients (Cluster One) showed low frequencies of impairments that appeared to be scattered across cognitive abilities, consistent with the classic pathological view of a widely distributed multifocal inflammatory disorder of the CNS. The other clusters had high frequencies, but, depending on the cluster, those high frequencies were on different tests. This indicated groups of patients with more severe problems in certain areas (e.g. verbal memory or processing speed).

Table 9. Frequencies (%) of patients failing each test in each cluster at phase I

a) Phase I, Complete cases (n=61). Ten-cluster solution

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
10-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	35	0	6	14	6	9	6	6	3	0	3	29	9	3	3	6	31
C2	2	50	0	0	0	0	**100**	0	50	0	50	0	0	0	0	**100**	0
C3	2	0	0	**100**	50	0	**100**	0	0	**100**	0	0	50	0	0	0	50
C4	7	**100**	**100**	**71**	43	0	43	0	0	0	14	57	0	29	0	0	14
C5	5	**80**	40	**100**	0	40	20	0	20	0	20	0	**100**	0	0	20	40
C6	1	0	0	0	0	0	0	0	**100**	**100**	**100**	0	0	0	0	**100**	**100**
C7	4	**75**	**100**	50	0	0	**100**	**75**	**100**	0	50	25	0	0	0	0	50
C8	2	50	0	50	0	0	**100**	**100**	**100**	0	0	**100**	0	0	50	**100**	50
C9	2	50	0	50	0	**100**	0	0	50	0	50	0	0	**100**	**100**	**100**	0
C10	1	**100**	**100**	**100**	**100**	**100**	**100**	0	**100**	**100**	0	0	**100**	**100**	**100**	0	0
Average																	
C2-10	26	**69**	**54**	**65**	19	19	**58**	19	42	15	27	27	27	19	15	31	31

b) Phase I, Imputation 1 (n=82). Six-cluster solution

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
6-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	48	8	2	15	4	8	10	6	13	2	13	21	13	6	4	15	38
C2	27	78	78	74	15	7	67	7	26	0	15	52	15	15	4	19	22
C3	2	50	0	100	50	0	100	0	0	100	0	0	50	0	0	0	50
C4	2	50	0	50	0	100	0	0	50	0	50	0	0	100	100	100	0
C5	2	100	100	100	100	100	100	50	100	100	0	50	50	50	100	0	0
C6	1	100	100	100	100	0	100	0	0	0	0	100	100	100	100	0	100
Average																	
C2-6	34	**77**	**71**	**76**	24	17	**68**	9	29	12	15	47	21	24	18	21	23

c) Phase I, Imputation 2 (n=82). Thirteen-cluster solution

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
13-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	41	0	2	15	5	7	10	5	5	0	7	27	5	10	0	2	29
C2	4	25	0	0	0	0	100	25	50	0	25	25	0	0	25	100	0
C3	13	92	92	77	31	0	38	0	8	0	8	69	8	15	8	23	15
C4	3	67	67	100	0	67	33	0	0	0	0	0	100	0	0	0	0
C5	6	67	83	33	0	0	100	0	50	0	50	0	0	33	0	0	50
C6	4	75	0	75	0	0	25	0	25	0	25	25	100	0	25	50	100
C7	2	50	0	100	50	0	100	0	0	100	0	0	50	0	0	0	50
C8	1	0	0	0	0	0	0	0	100	100	100	0	0	0	0	100	100
C9	3	100	67	100	0	33	100	100	100	0	0	67	0	0	0	33	33
C10	2	50	0	50	0	100	0	0	50	0	50	0	0	100	100	100	0
C11	1	100	100	100	100	100	100	0	100	100	0	0	100	100	100	0	0
C12	1	100	100	100	100	100	100	0	100	0	100	100	100	0	0	100	0
C13	1	100	100	100	100	0	0	0	0	0	100	100	100	100	0	0	100
Average																	
C2-13	41	**75**	**60**	**67**	20	18	**60**	10	35	10	25	37	30	20	15	35	32

d) Phase I, Imputation 3 (n=82). Thirteen-cluster solution

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
13-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	41	0	2	12	5	10	7	7	5	0	7	20	7	7	0	7	29
C2	16	100	81	56	19	13	63	0	19	0	13	31	19	19	0	6	6
C3	8	50	50	100	0	0	75	25	38	0	0	100	13	0	0	63	13
C4	2	50	0	100	50	0	100	0	0	100	0	0	50	0	0	0	50
C5	3	100	0	67	0	0	0	0	33	0	33	0	100	0	33	33	100
C6	3	33	100	67	0	0	100	33	67	0	67	0	0	33	0	0	100
C7	1	0	0	0	0	0	0	0	100	100	100	0	0	0	0	100	100
C8	1	100	100	0	0	0	100	0	0	0	100	100	0	100	0	100	0
C9	1	0	0	0	0	0	100	100	100	0	0	100	0	0	100	100	0
C10	2	50	0	50	0	100	0	0	50	0	50	0	0	100	100	100	0
C11	1	100	100	100	100	100	100	0	100	100	0	0	100	100	100	0	0
C12	1	100	100	100	100	100	100	0	100	100	100	100	100	0	0	100	0
C13	2	100	100	100	100	0	0	0	0	50	0	100	50	0	100	0	50
Average																	
C2-13	41	**78**	**62**	**70**	20	15	**63**	10	35	15	23	45	28	20	18	33	28

e) Phase I, Imputation 4 (n=82). Fifteen-cluster solution

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
15-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	38	0	3	8	5	8	5	8	5	0	3	21	5	8	0	3	32
C2	3	0	0	0	0	0	67	0	33	0	67	67	0	0	67	100	0
C3	4	75	0	75	0	0	0	0	25	0	50	0	75	0	25	25	100
C4	6	100	100	100	83	0	83	0	0	0	0	33	17	0	17	0	33
C5	3	67	67	100	0	67	33	0	0	0	0	0	100	0	0	0	0
C6	5	100	80	40	0	0	0	0	0	0	20	80	20	40	0	0	0
C7	6	33	50	100	0	0	67	0	17	0	0	100	17	0	0	67	0
C8	4	50	100	50	0	0	100	0	50	0	75	0	0	25	0	0	100
C9	4	100	50	25	0	25	100	0	75	0	0	0	0	25	0	25	0
C10	2	50	0	100	50	0	100	0	0	100	0	0	50	0	0	0	50
C11	1	0	0	0	0	0	0	0	100	100	100	0	0	0	0	100	100
C12	2	100	50	100	0	0	100	100	100	0	0	100	0	0	0	50	50
C13	2	50	0	50	0	100	0	0	50	0	50	0	0	100	100	100	0
C14	1	100	100	100	100	100	100	0	100	100	0	0	100	100	100	0	0
C15	1	100	100	100	100	100	100	0	100	0	100	100	100	0	0	100	0
Average																	
C2-15	44	**70**	**56**	**70**	19	16	**61**	5	33	9	26	40	28	16	16	33	30

f) Phase I, Imputation 5 (n=82). Fourteen-cluster solution

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
14-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	40	0	3	3	5	10	5	5	5	0	8	23	8	10	0	3	33
C2	4	50	0	0	0	0	100	25	25	0	25	0	0	25	0	75	0
C3	5	20	40	100	0	0	80	0	0	0	0	100	20	0	0	60	0
C4	1	0	0	0	0	0	100	0	100	0	0	100	0	0	100	100	0
C5	3	100	0	67	0	0	0	0	33	0	33	0	100	0	33	33	100
C6	2	50	0	100	50	0	100	0	0	100	0	0	50	0	0	0	50
C7	8	100	100	88	38	25	100	25	38	0	0	25	13	0	0	0	13
C8	7	100	86	57	14	0	0	0	0	0	14	71	29	29	14	0	0
C9	4	50	100	50	0	0	100	0	50	0	75	0	0	25	0	0	100
C10	1	0	0	0	0	0	0	0	100	100	100	0	0	0	0	100	100
C11	2	100	50	100	0	0	50	50	100	0	0	100	0	0	0	100	100
C12	2	50	0	50	0	100	0	0	50	0	50	0	0	100	100	100	0
C13	1	100	100	100	100	100	100	0	100	100	0	0	100	100	100	0	0
C14	2	100	100	100	100	50	100	0	50	100	100	100	50	0	0	0	50
Average																	
C2-14	42	**73**	**59**	**68**	20	15	**66**	10	34	15	24	41	25	17	15	32	32

Table 9 shows that at phase I around half of pwMS had multifocal impairments, and half had predominantly verbal memory and processing speed impairments (SDMT only), accompanied with or without additional deficits on VOSP6. This was consistent across all imputations and the complete case analysis.

Table 9 shows dissociation between scattered impairments in Cluster One, indicative of widely distributed multifocal inflammatory disorder of the CNS; and other clusters with more pronounced damage. Some of the clusters grouped together patients with more damage on some tests but not the others, however, that distinction was unclear. Only once the remaining clusters that grouped together around half of patients and phase I were merged together, a pattern of deficits emerged with predominantly verbal memory and processing speed impairments (SDMT only), accompanied with or without additional deficits on VOSP6. This was consistent across all imputations and the complete case analysis.

Table 10. Comparison of MS types between Cluster One and the remaining clusters at phase I (n=82)

	'Multifocal' cluster (C1)		**All other clusters**	
	Size (n)	**Types**	**Size (n)**	**Types**
Complete case analysis (n=61)	35	26% PPMS, 23% SPMS, 51% RRMS	26	27% PPMS, 32% SPMS, 41% RRMS
Imputation 1	48	24% PPMS, 29% SPMS, 47% RRMS	34	22% PPMS, 42% SPMS, 36% RRMS
Imputation 2	41	27% PPMS, 19% SPMS, 54% RRMS	41	20% PPMS, 49% SPMS, 31% RRMS
Imputation 3	41	27% PPMS, 22% SPMS, 51% RRMS	41	20% PPMS, 46% SPMS, 34% RRMS
Imputation 4	38	29% PPMS, 18% SPMS, 53% RRMS	44	18% PPMS, 48% SPMS, 34% RRMS
Imputation 5	40	28% PPMS, 18% SPMS, 54% RRMS	42	19% PPMS, 50% SPMS, 31% RRMS

Table 10 shows that at phase I the majority of participants in Cluster One were RRMS, and in other clusters the majority of pwMS were SPMS. These differences in MS type prevalence were more pronounced in the MI datasets.

Abbreviations: PPMS – Primary Progressive MS, SPMS – Secondary Progressive MS, RRMS – Relapsing-Remitting MS

Therefore in all further analyses it was considered to merge the remaining clusters and compare them against the Cluster One. As it can be seen from Table 10, the two halves of participants had differed in the prevalence of MS subtypes. The 'Multifocal' cluster, or Cluster One, was largely comprised of the RRMS patients, and the other clusters with more pronounced verbal memory and processing speed impairments, were comprised largely of SPMS patients.

3.5.2. Subject-clusters in the phase II dataset

After having examined the subject-clusters in phase I, the next step was to investigate the dimensionality of the phase II cohort, and whether there had been any changes in the groupings of patients based on their cognitive profiles.

Table 11. Determination of the optimal number of clusters in the hierarchical cluster analysis of patient data at phase II (n=82)

	Caliński-Harabasz pseudo-F					
Number of Clusters	**Complete cases (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
2	**23.01**	11.43	**23.21**	8.52	2.53	**11.44**
3	14.06	**16.34**	13.07	5.44	**13.38**	6.96
4	12.54	12.34	11.02	**11.24**	11.30	11.08
5	10.7	10.94	9.41	9.06	9.59	9.63
6	9.39	9.28	8.18	8.08	8.31	9.85
7	-	9.04	7.97	7.84	8.40	8.92
8	7.65	8.44	8.15	7.39	8.55	8.13
9	7.22	8.18	7.87	7.33	8.17	8.35
10	8.19	7.51	7.47	7.71	7.72	7.83
11	8.04	7.04	7.00	7.40	7.17	7.38
12	-	6.82	6.58	7.00	6.88	6.93
13	7.23	6.97	6.31	6.63	6.54	6.65
14	7.16	6.78	6.04	6.40	6.23	6.54
15	-	6.59	5.81	6.15	5.98	7.37

Table 11 shows that the complete cases and all five MI datasets had yielded non-different numbers of cluster solutions, ranging from two to four clusters.

According to the Caliński-Harabasz cluster stopping criterion, all analyses suggest mainly the two or three cluster solutions in the phase II datasets. This indicated that the clustering solution at phase II was more stable than at phase I, as all datasets (complete cases and all MIs) yielded similar grouping solutions. Therefore it could be considered that the results from the complete cases analysis were more robust at phase II than they were at phase I.

Again, as can be seen from the Table 12 and the dendrograms (Figure 6), in each clustering solution the clusters were unequal in their sizes. All solutions suggested one large cluster that contained the majority of patients, and then one or more other smaller clusters that had much less participants in them (1 to 26%).

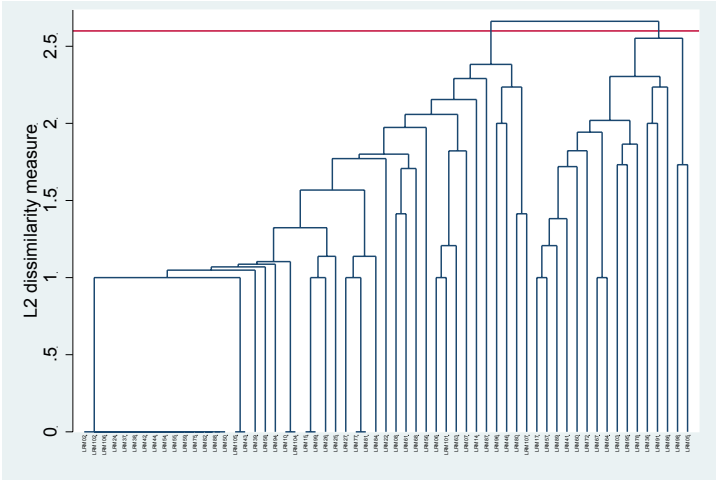
Table 12. Number of individuals in each subject-cluster solution at phase II (n=82)

Cluster	Complete cases (n=61)	Imputation 1	Imputation 2	Imputation 3	Imputation 4	Imputation 5
1	45 (74%)	61 (74%)	61 (75%)	61 (74%)	60 (73%)	76 (93%)
2	16 (26%)	16 (20%)	21 (25%)	16 (20%)	21 (26%)	6 (7%)
3		5 (6%)		1 (1%)	1 (1%)	
4				4 (5%)		

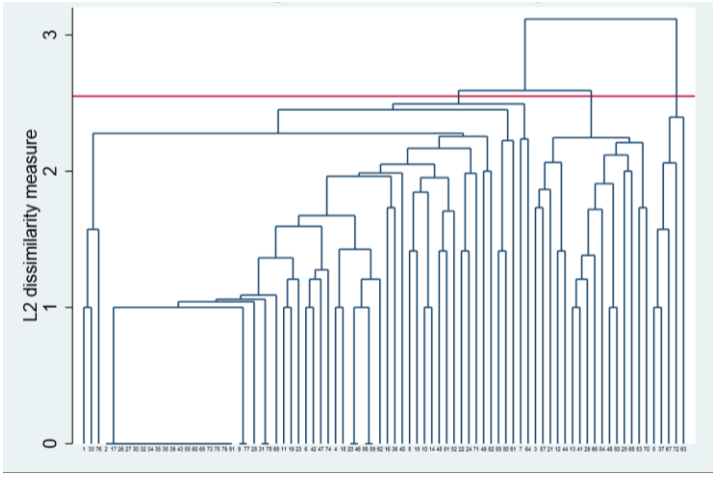
Table 12 shows that in all subject-cluster solutions there was a one larger cluster with three quarters of pwMS and the remaining quarter was distributed among one to three small clusters. The results from all six datasets were similar, except for imputation 5 that has offered a much more homogeneous solution for the dataset.

Figure 6. Dendrograms of hierarchical cluster analysis using average linkage model of patient data at phase II. The horizontal line is shown to represent where the optimal cluster cut falls in each analysis

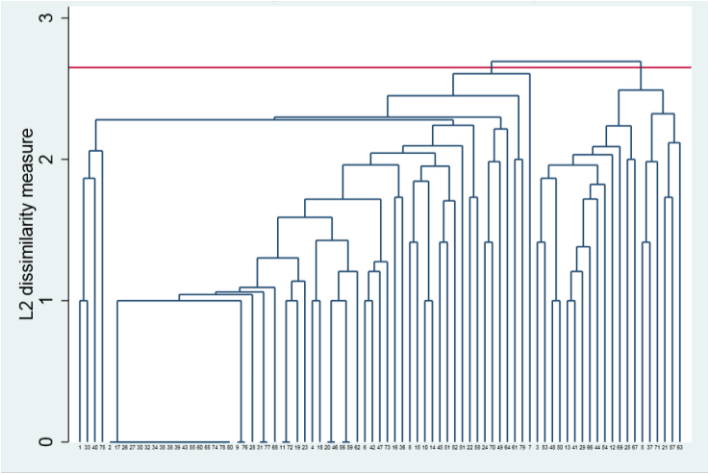
a) Phase II, Complete cases (n=61). Two-cluster solution



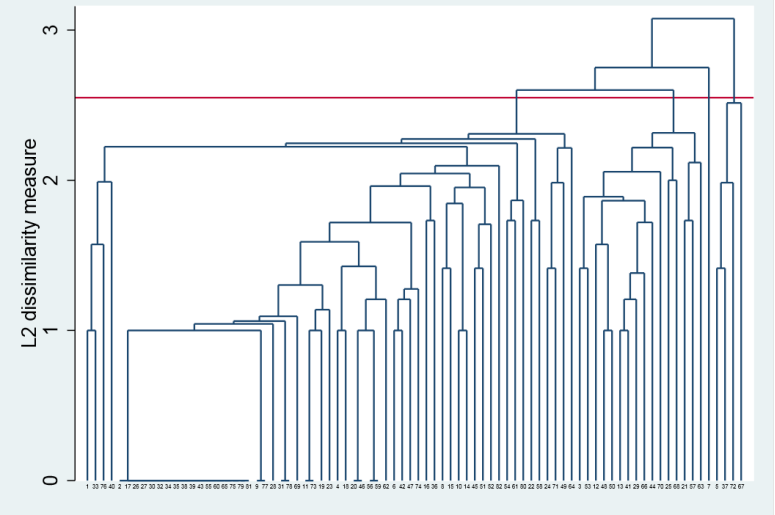
b) Phase II, Imputation 1 (n=82). Three-cluster solution



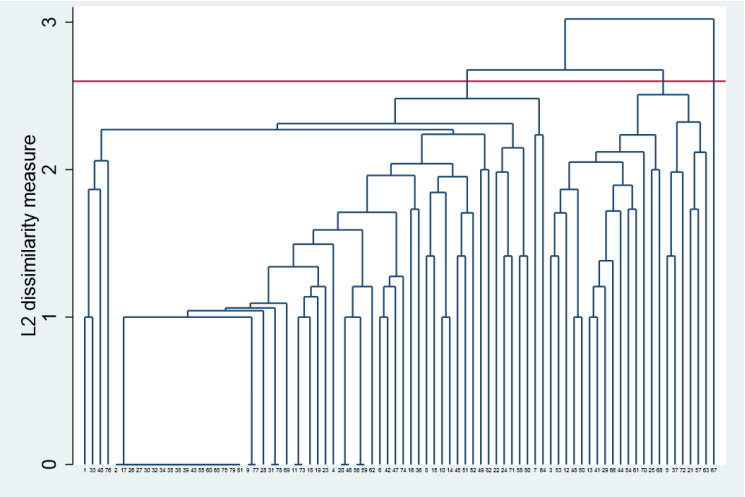
c) Phase II, Imputation 2 (n=82). Two-cluster solution



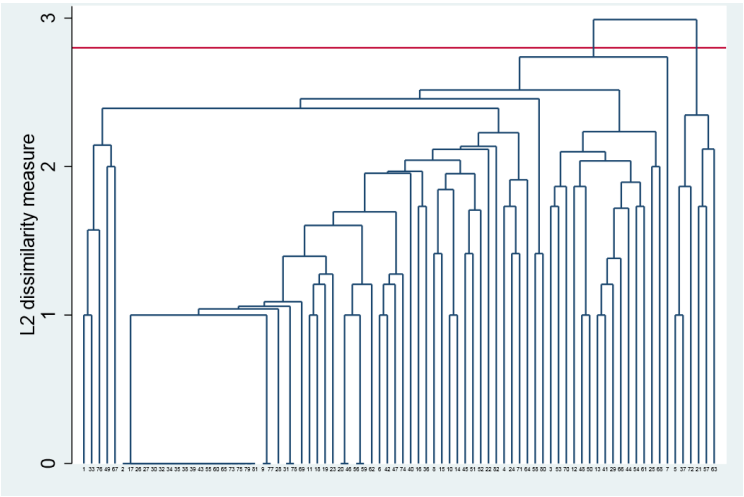
d) Phase II, Imputation 3 (n=82). Four-cluster solution



e) Phase II, Imputation 4 (n=82). Three-cluster solution



f) Phase II, Imputation 5 (n=82). Two-cluster solution



After determining the most appropriate numbers of clusters for each dataset, the next step was to explore the cognitive profiles of the patients in the large subject-cluster, and then to compare them to the profiles of the patients in the smaller subject-clusters.

Table 13. Frequencies (%) of patients failing each test in each subject-cluster at phase II

a) Phase II Complete Case analysis (n=61)

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Langu-age	
2-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	45	11	4	13	9	16	27	31	7	9	9	4	11	4	16	9	4
C2	16	**100**	**88**	**88**	31	13	**100**	**88**	6	31	44	**50**	38	25	31	25	0

b) Phase II Imputation 1 (n=82)

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Langu-age	
3-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	61	16	5	16	10	15	36	36	10	11	16	7	15	8	16	10	3
C2	16	100	88	88	38	19	100	88	19	19	38	44	31	25	13	13	0
C3	5	100	80	100	80	60	100	80	40	100	80	80	80	40	100	100	0
Average C2-3	21	**100**	**86**	**91**	48	29	**100**	**86**	24	38	48	53	43	29	34	34	0

c) Phase II Imputation 2 (n=82)

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Langu-age	
2-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	61	16	5	16	10	15	36	38	10	11	16	7	15	8	18	10	3
C2	21	**100**	**86**	**90**	43	29	**100**	**90**	24	38	48	43	43	29	33	33	0

d) Phase II Imputation 3 (n=82)

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Langu-age	
4-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	61	18	5	16	8	11	38	39	10	10	16	8	13	8	15	8	3
C2	16	100	94	94	31	13	100	94	19	25	38	31	38	31	19	19	0
C3	1	0	0	0	100	100	0	0	0	100	0	0	100	0	100	100	0
C4	4	100	75	75	75	50	100	75	50	100	100	75	75	25	100	100	0
Average																	
C2-4	21	**95**	**86**	**86**	43	24	**95**	**86**	24	43	48	38	48	28	38	38	0

e) Phase II Imputation 4 (n=82)

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
3-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	60	17	5	15	10	13	35	37	10	12	17	5	15	7	18	10	3
C2	21	95	86	90	43	19	100	90	19	33	43	43	38	29	29	29	0
C3	1	0	0	100	0	0	100	100	00	100	100	100	100	100	100	100	0
Average																	
C2-3	22	**91**	**82**	**90**	41	18	**100**	**90**	18	36	46	46	41	32	32	32	0

e) Phase II Imputation 5 (n=82)

		Verbal memory			Visuo-spatial memory		Processing speed			Visuoperceptual						Language	
2-cluster	Size (N)	SRTL	SRTC	SRTD	SPART	SPARTD	SDMT	PASAT	WLGT	VOSP2	VOSP3	VOSP6	VOSP8	BORB2	BORB5	TROG	BC.Index
C1	76	32	20	29	12	14	49	49	13	14	22	13	17	11	17	9	3
C2	6	**100**	**100**	**100**	83	**50**	**100**	**83**	17	67	**50**	**67**	**83**	33	**67**	**100**	0

Table 13 shows that at phase II the participants again could be split into two groups based on their cognitive profiles. The first larger cluster with three quarters of patients showed multifocal widely distributed deficits. However, the remaining patients in other clusters exhibited a more progressive cognitive impairment, with deficits in verbal memory and processing speed being most predominant.

As it can be noted from the Table 13, at phase II the patients could be grouped into two distinct groups. The larger cluster that contained approximately three quarters of the patients was similar to the larger group at phase I – it included patients with widely distributed impairments without a coherent pattern. However, what was different at phase II from phase I, was that at phase II the second largest cluster could be clearly defined by problems with verbal memory (SRT) and information processing speed (SDMT and PASAT) with or without additional problems. It appears that in the three years between phase I and phase II assessments the sample had become more homogeneous. It had formed into two groups of patients, those with mild scattered impairments, and those with a more progressive pattern. This profile was consistent among all datasets – the complete cases and all five MIs.

Although in phase I the more progressive clusters grouped together deficits on many domains, at phase II the more progressive patients were more homogeneous, with major problems settling on verbal memory and information processing speed.

Table 14. MS types in each subject-cluster at phase II (n=82)

	'Multifocal' cluster (C1)		'Memory/proc speed' cluster (all other clusters)	
	Size (n)	**MS Types**	**Size (n)**	**MS Types**
Complete case analysis (n=61)	45	29% PPMS, 16% SPMS, 55% RRMS	16	19% PPMS, 56% SPMS, 25% RRMS
Imputation 1	61	26% PPMS, 21% SPMS, 52% RRMS	21	14% PPMS, 71% SPMS, 14% RRMS
Imputation 2	61	26% PPMS, 21% SPMS, 52% RRMS	21	14% PPMS, 71% SPMS, 14% RRMS
Imputation 3	61	25% PPMS, 21% SPMS, 54% RRMS	21	19% PPMS, 71% SPMS, 10% RRMS
Imputation 4	60	27% PPMS, 20% SPMS, 53% RRMS	22	14% PPMS, 71% SPMS, 14% RRMS
Imputation 5	76	25% PPMS, 29% SPMS, 46% RRMS	6	100% SPMS

Table 14 shows that the more impaired on memory and processing speed were mainly SPMS patients

Abbreviations: PPMS – Primary Progressive MS, SPMS – Secondary Progressive MS, RRMS – Relapsing-Remitting MS

As it can be seen from Table 14, the more impaired memory and processing speed group were mainly comprised of SPMS patients, with their cognitive profile more indicative of progressive disease course.

3.5.3. Subject-clusters of the trajectory of longitudinal change

As it can be seen from the analyses in section 3.5.1. and section 3.5.2. earlier in this appendix, the cluster analyses have yielded slightly different results when employed at phase I and at phase II. This was consistent across all datasets (complete cases and the five MIs). This has allowed to assume that there have been time-related changes in the dimensionality of the study population. To address this in the following section the temporal changes in MS-related cognitive deficits were examined by

using the cluster analysis to group the patients based on the trajectory of their longitudinal changes. This was carried out to identify individuals whose individual cognitive deficits have progressed, remained the same, or improved over the period of three years.

As it can be seen from Table 15, each dataset (complete cases and the five MIs) has yielded a different grouping solution. Depending on the dataset, the numbers of groups have varied between four and fourteen. This indicated that the subject-clustering solutions provided by the longitudinal analysis were not robust, and should be interpreted with caution.

Moreover, the Calinski-Harabasz pseudo-F indexes were very similar for all subject-cluster solutions in all datasets, most of them were around 3 or 4 in values; when in comparison to phase II, where these indexes were 5 to 23. Since the differences between Calinski-Harabasz Pseudo-F indexes were so minimal, this has again showed that the results from the subject-clustering solutions of longitudinal change were not robust and shouldn't be used to draw conclusions.

Table 15. Determination of the optimal number of clusters in the hierarchical cluster analysis of longitudinal change in their performance between phases I and II (n=82)

	Caliński-Harabasz pseudo-F					
Number of Clusters	**Complete cases (n=61)**	**Imputation 1**	**Imputation 2**	**Imputation 3**	**Imputation 4**	**Imputation 5**
2	2.4	2.98	2.28	2.63	3.59	2.39
3	4.06	3.48	3.13	2.37	2.90	3.19
4	**4.23**	**4.42**	3.52	2.96	3.37	3.49
5	3.66	3.74	3.21	3.56	3.88	3.63
6	4.22	3.99	2.91	3.60	3.88	3.59
7	3.84	3.88	3.66	**4.14**	3.63	**3.98**
8	3.63	3.66	3.63	3.80	3.38	3.69
9	3.83	4.08	3.47	3.59	3.68	3.53
10	3.61	3.84	3.70	3.80	3.47	3.68
11	3.65	3.92	3.67	3.76	3.84	3.50
12	3.52	3.79	3.73	3.59	**3.92**	3.34
13	3.46	3.61	3.55	3.96	3.73	3.56
14	3.41	3.48	**3.78**	3.64	3.62	3.41
15	-	3.73	3.69	3.58	3.86	3.67

Table 15 shows that there was no consensus among the datasets for the optimum number of subject-clusters. Depending on a dataset (complete cases or the five MIs) the number of clusters ranged from four to fourteen.

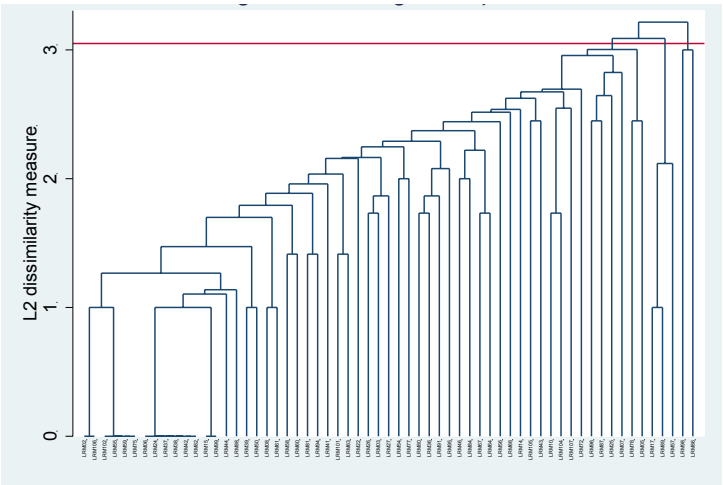
Table 16. Number of individuals in each subject-cluster solution at phase II

Cluster	Complete cases (n=61)	Imputation 1	Imputation 2	Imputation 3	Imputation 4	Imputation 5
C1	54 (89%)	72 (88%)	55 (67%)	67 (82%)	58 (71%)	67 (82%)
C2	3 (5%)	6 (7%)	4 (5%)	5 (6%)	3 (4%)	3 (4%)
C3	3 (5%)	3 (4%)	3 (4%)	2 (2%)	5 (6%)	2 (2%)
C4	1 (2%)	1 (1%)	2 (3%)	3 (4%)	1 (1%)	3 (4%)
C5			4 (5%)	3 (4%)	3 (4%)	3 (4%)
C6			1 (1%)	1 (1%)	1 (1%)	3 (4%)
C7			3 (4%)	1 (1%)	2 (2%)	1 (1%)
C8			3 (4%)		3 (4%)	
C9			1 (1%)		2 (2%)	
C10			1 (1%)		1 (1%)	
C11			1 (1%)		1 (1%)	
C12			1 (1%)		2 (2%)	
C13			2 (2%)			
C14			1 (1%)			

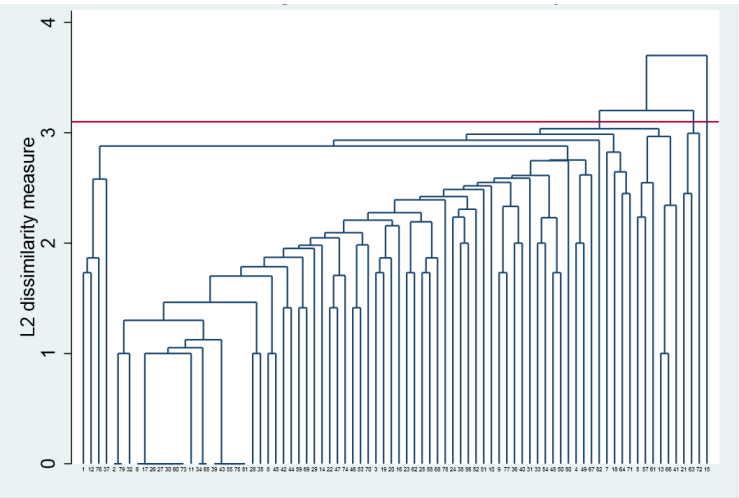
Table 16 showed that in all datasets Cluster One had grouped the majority of participants, with minor numbers of patients in the remaining subject-clusters

Figure 7. Dendrograms of hierarchical cluster analysis model of longitudinal trajectories of change in patient performance on cognitive tests between phases I and II (n=82).

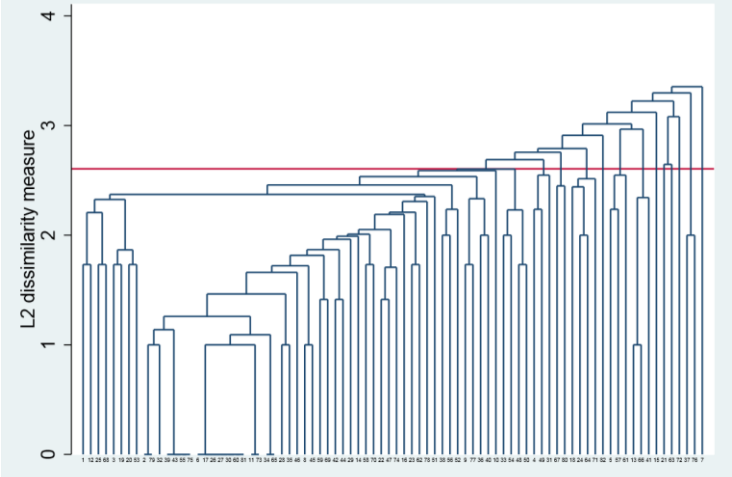
a) Longitudinal Change, Complete Case Analysis (n=61). Four-cluster solution



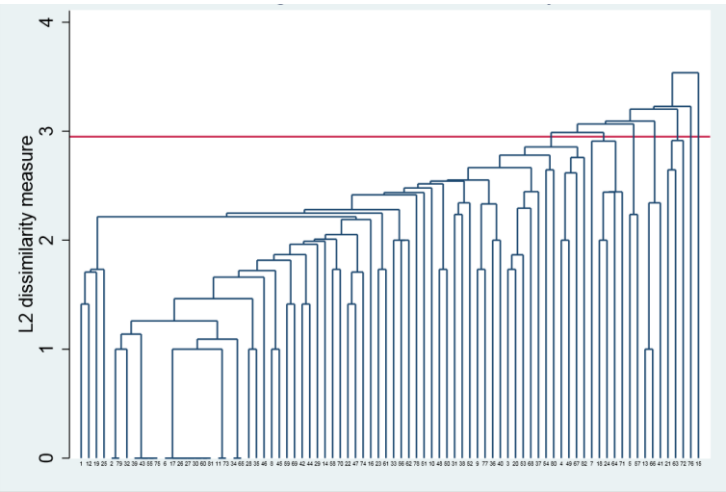
b) Longitudinal Change, Imputation 1. Four-cluster solution



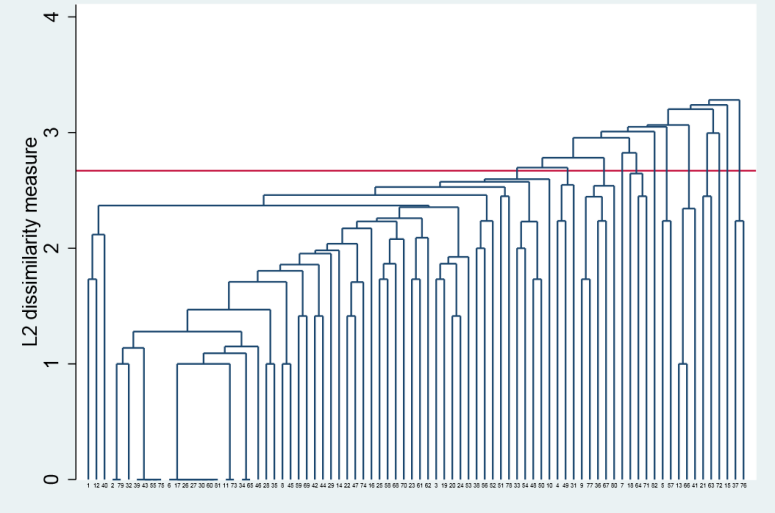
c) Longitudinal Change, Imputation 2. Fourteen-cluster solution



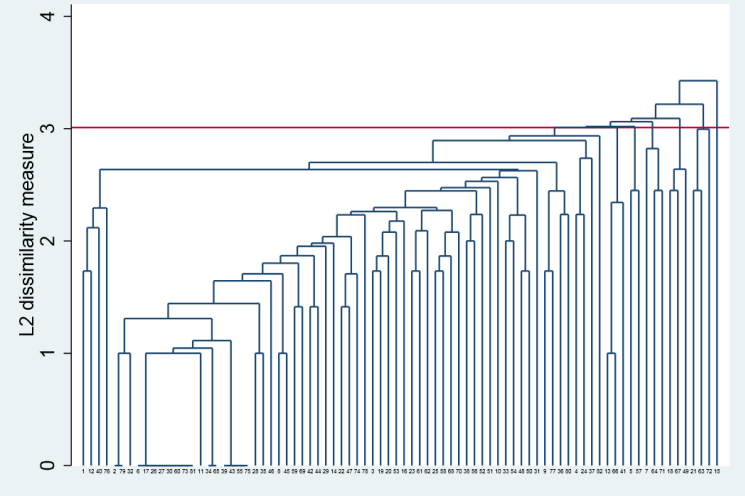
d) Longitudinal Change, Imputation 3. Seven-cluster solution



e) Longitudinal Change, Imputation 4. Twelve-cluster solution



f) Longitudinal Change, Imputation 5. Seven-cluster solution



In all datasets (complete cases and five MIs) there was a clear one large cluster with the majority of patients. This large subject-cluster has grouped together the patients that had mostly unchanged, with similar frequencies of resolved impairments and new impairments on all cognitive tests (Table 17). However, in the smaller clusters a shift could be observed towards acquiring more new impairments and much less resolved impairments, indicative of a faster and more progressive cognitive deterioration. The emerging impairments were most evident on the verbal memory (SRT) and information processing speed items (SPART and PASAT), and could also be seen, but to a lesser extent on some of the visuoperceptual items. This pattern was consistent across all datasets of longitudinal change. Therefore to show this contrast in further analyses all the smaller clusters were merged into one in order to contrast them to the largest C1 in each dataset. After investigating the patterns of longitudinal changes of the patients in the many small clusters, no clear differences were found. This again has supported the initial complete cases analysis performed in Chapter Five.

Table 17. Neuropsychological characteristics for the subject-cluster solutions for longitudinal change for each dataset. Frequencies (%) of three different trajectories of change on the cognitive tests are presented

a) Complete Case Analysis (n=61). Four-cluster Solution

Test	Cluster One (n=54)			Other Clusters (n=7)		
	Resolved Impairment	Unchanged	New impairment	Resolved Impairment	Unchanged	New impairment
SRTL	13	78	9	0	29	71
SRTC	11	87	2	0	29	71
SRTD	17	79	4	0	29	71
SPART	6	88	6	0	71	29
SPARTD	7	84	9	0	100	0
SDMT	2	81	17	0	57	43
PASAT	0	69	31	0	43	57
WLGT	15	79	6	43	57	0
VOSP2	4	87	9	0	71	29
VOSP3	7	84	9	0	71	29
VOSP6	24	70	6	14	29	57
VOSP8	13	74	13	14	43	43
BORB2	6	90	4	14	57	29
BORB5	0	87	13	0	71	14
TROG	9	91	0	0	57	43
BC. Index	33	63	4	0	86	0

b) Imputation 1 (n=82). Four-cluster Solution

Test	Cluster One (n=72)			Other Clusters (n=10)		
	Resolved Impairment	**Unchanged**	**New impairment**	**Resolved Impairment**	**Unchanged**	**New impairment**
SRTL	13	79	8	10	40	50
SRTC	15	82	3	10	30	60
SRTD	17	81	3	0	40	60
SPART	6	83	11	10	60	30
SPARTD	7	82	11	0	80	20
SDMT	1	79	19	10	60	30
PASAT	0	**65**	**35**	0	**10**	**90**
WLGT	12	78	10	30	70	0
VOSP2	3	83	14	0	**80**	**20**
VOSP3	7	78	15	0	**70**	**30**
VOSP6	22	74	4	30	20	50
VOSP8	11	71	18	20	60	20
BORB2	8	85	7	20	50	30
BORB5	0	86	14	20	70	10
TROG	10	87	3	0	**60**	**40**
BC. Index	33	64	3	20	80	0

c) Imputation 2 (n=82). Fourteen-cluster solution.

Test	Cluster One (n=55)			Other Clusters (n=27)		
	Resolved Impairment	**Unchanged**	**New impairment**	**Resolved Impairment**	**Unchanged**	**New impairment**
SRTL	9	86	6	18	52	30
SRTC	9	87	4	26	52	22
SRTD	11	85	4	26	52	22
SPART	7	86	7	4	74	22
SPARTD	7	82	11	4	81	15
SDMT	2	82	16	0	**74**	**26**
PASAT	0	**75**	**25**	0	**18**	**82**
WLGT	5	84	11	33	63	4
VOSP2	4	89	7	0	**67**	**33**
VOSP3	0	85	15	18	67	15
VOSP6	24	73	4	30	48	22
VOSP8	13	82	5	11	48	41
BORB2	4	91	5	26	56	18
BORB5	0	**89**	**11**	4	70	26
TROG	**7**	**93**	0	11	70	19
BC. Index	25	73	2	41	55	4

d) Imputation 3 (n=82). Seven-cluster solution

Test	Cluster One (n=67)			Other Clusters (n=15)		
	Resolved Impairment	**Unchanged**	**New impairment**	**Resolved Impairment**	**Unchanged**	**New impairment**
SRTL	10	81	9	27	40	33
SRTC	15	82	3	20	40	40
SRTD	18	78	4	7	60	33
SPART	7	85	7	7	60	33
SPARTD	9	82	9	0	**87**	**13**
SDMT	1	81	18	0	**73**	**27**
PASAT	0	**66**	**34**	0	**20**	**80**
WLGT	10	81	9	33	60	7
VOSP2	3	87	10	7	60	33
VOSP3	4	81	15	13	67	20
VOSP6	22	72	6	40	33	27
VOSP8	12	78	10	13	40	47
BORB2	4	88	8	33	47	20
BORB5	0	**87**	**13**	13	67	20
TROG	**7**	**91**	2	20	53	27
BC. Index	30	69	1	20	73	7

e) Imputation 4 (n=82). Twelve-cluster solution

Test	Cluster One (n=58)			Other Clusters (n=24)		
	Resolved Impairment	**Unchanged**	**New impairment**	**Resolved Impairment**	**Unchanged**	**New impairment**
SRTL	5	88	7	33	38	29
SRTC	5	91	3	38	38	25
SRTD	7	88	5	33	46	21
SPART	5	86	9	12	63	25
SPARTD	5	86	9	12	75	12
SDMT	2	77	21	4	75	21
PASAT	0	**64**	**36**	0	**33**	**67**
WLGT	10	79	10	25	71	4
VOSP2	3	88	9	0	**67**	**33**
VOSP3	3	83	14	12	67	21
VOSP6	24	69	7	25	58	17
VOSP8	12	81	7	12	46	42
BORB2	3	93	3	21	54	25
BORB5	0	**90**	**10**	4	71	25
TROG	**7**	**93**	0	12	67	21
BC.Index	31	67	2	29	67	4

f) Imputation 5 (n=82). Seven-cluster solution

	Cluster One (n=67)			Other Clusters (n=15)		
Test	**Resolved Impairment**	**Unchanged**	**New impairment**	**Resolved Impairment**	**Unchanged**	**New impairment**
SRTL	10	81	9	27	40	33
SRTC	12	85	3	27	33	40
SRTD	15	81	4	20	47	33
SPART	6	85	9	13	6	27
SPARTD	6	81	13	13	80	7
SDMT	**2**	**79**	**19**	**7**	**73**	**20**
PASAT	0	**64**	**36**	0	**20**	**80**
WLGT	10	79	10	33	67	0
VOSP2	3	85	12	7	67	27
VOSP3	5	82	13	13	67	20
VOSP6	22	72	6	33	40	27
VOSP8	12	78	10	**13**	**33**	**53**
BORB2	7	87	6	20	60	20
BORB5	0	**87**	**14**	7	73	20
TROG	**7**	**91**	1	**3**	**53**	**33**
BC.Index	36	63	1	13	80	7

Table 17 showed that in all longitudinal datasets (complete cases and five MI) the profile for change of individuals in the large cluster was sporadic. However, the other participants that fell into the remaining three to thirteen subject-clusters, exhibited a more progressive accumulation of new impairments, mainly on the information processing speed items.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

Table 18. Clinical characteristics in each subject-cluster of longitudinal change (n=82)

	'Multifocal' change cluster (C1)		**'Progressive' cluster (all other clusters)**		**Difference**
	Size (n)	**Types**	**Size (n)**	**Types**	
Complete case analysis (n=61)	54	29% PPMS, 17% SPMS, 54% RRMS	7	100% SPMS	**X^2 = 18.89, p < 0.001**
Imputation 1	72	26% PPMS, 26% SPMS, 47% RRMS	10	90% SPMS, 10% RRMS	**X^2 = 15.89, p < 0.001**
Imputation 2	55	29% PPMS, 22% SPMS, 49% RRMS	27	11% PPMS, 59% SPMS, 23% RRMS	**X^2 = 11.57, p = 0.003**
Imputation 3	67	27% PPMS, 27% SPMS, 46% RRMS	15	7% PPMS, 67% SPMS, 27% RRMS	**X^2 = 8.95, p = 0.011**
Imputation 4	58	26% PPMS, 21% SPMS, 53% RRMS	24	17% PPMS, 67% SPMS, 17% RRMS	**X^2 = 16.51, p < 0.001**
Imputation 5	67	27% PPMS, 24% SPMS, 49% RRMS	15	7% PPMS, 80% SPMS, 13% RRMS	**X^2 = 17.17, p < 0.001**

Table 18 showed that the smaller clusters contained much more patients with SPMS
Note. The comparison of frequencies was done with Pearson Chi-Square test
Abbreviations: PPMS – Primary Progressive MS, SPMS – Secondary Progressive MS, RRMS – Relapsing-Remitting MS

After comparing the two groups of patients (large relatively unchanged Cluster One and the remaining clusters with pwMS that deteriorated), it was found that those who have deteriorated had predominantly SPMS course (Table 18). These individuals were also older, had longer disease duration, more neurological disability, higher levels of depression and disease impact; but none of those differences reached statistical significance. Therefore it can be concluded that the main factor which has caused progression of cognitive symptoms and new deficit acquirement was the MS course.

However, since the clustering solutions in this instance were not robust, the dimensionality of the datasets of longitudinal change should be interpreted with caution.

3.5.4. Dimensionality of the study population. Summary

In the second half of this appendix the dimensionality of the study cohort was investigated by using the cluster analysis method. It was found that the study population had a tendency to become more homogeneous over time, emphasizing the differences between the more progressive MS course.

At phase I around half of patients had multifocal impairments, and half exhibited a pattern of more pronounced verbal memory and visual information processing speed impairments. At phase II the more progressive patients had developed more deficits on the information processing speed tasks. From analysing the datasets of longitudinal change it can be seen that the newly emerging deficits were mainly in the information processing speed domain, and had affected predominantly the SPMS patients. However, since the clustering solutions hadn't been robust, these findings should be interpreted with caution.

3.6. Section summary

The analyses with multiple imputation (MI) presented in this appendix have supported the complete cases analyses performed in Chapter Five.

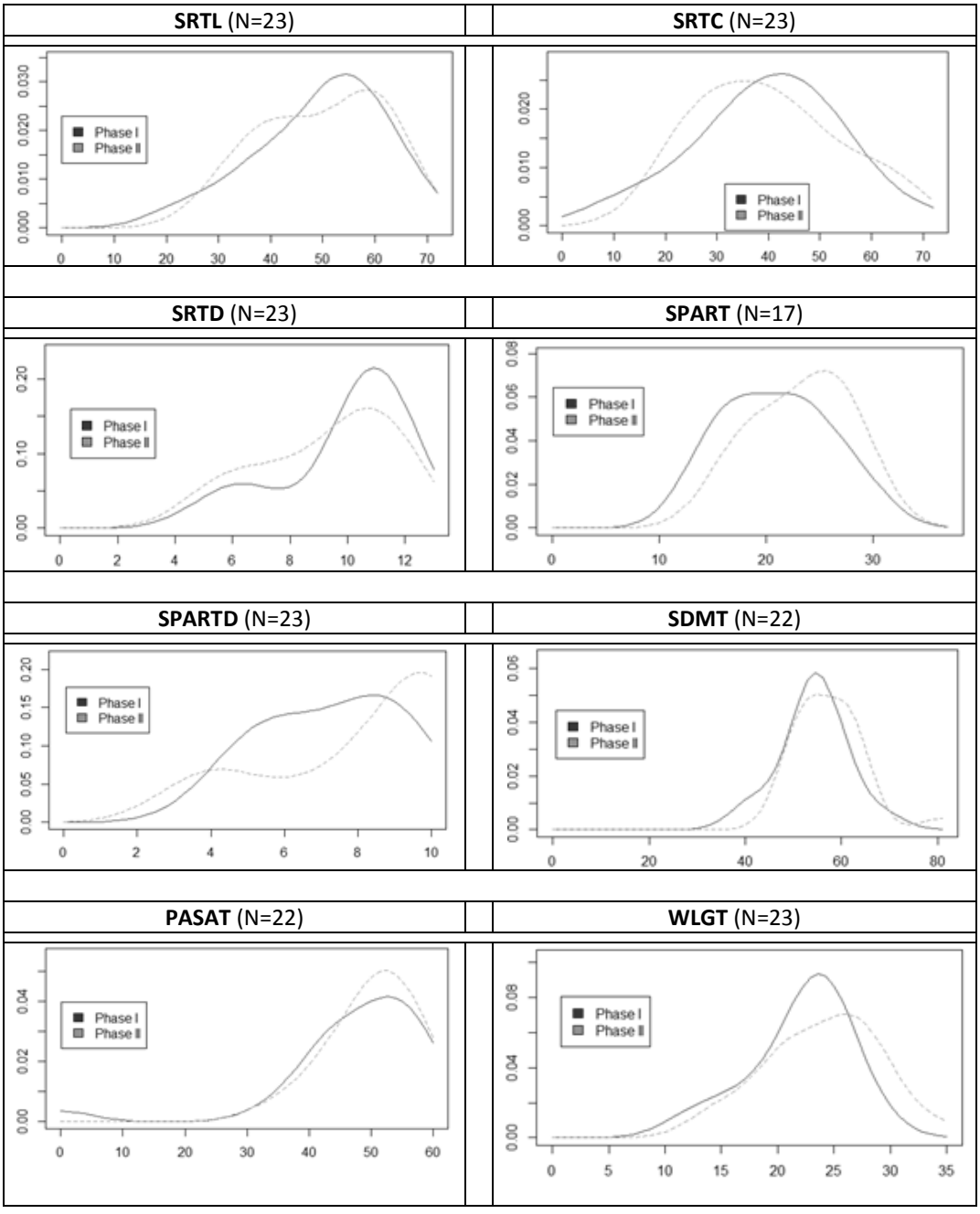
In this appendix it was found that both the analyses of the dimensionality of cognitive impairments in pwMS, and the analyses of the dimensionality of the study cohort, have yielded similar results for the complete cases and the MI datasets. This has shown that the findings presented in Chapter Five were robust and that the conclusions drawn about the dimensionality of cognitive impairments in MS had not been specific only to the subgroup of patients who had complete data.

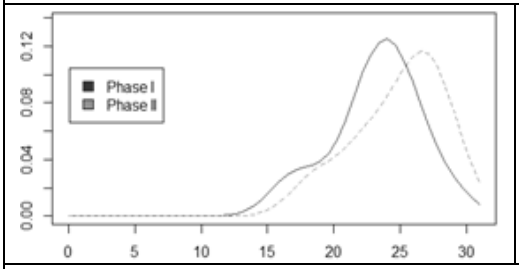
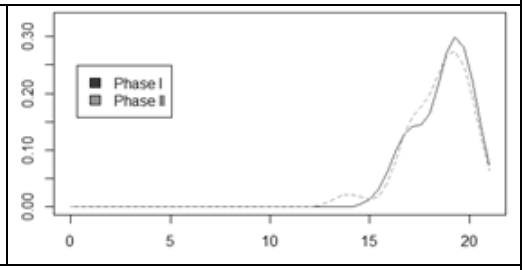
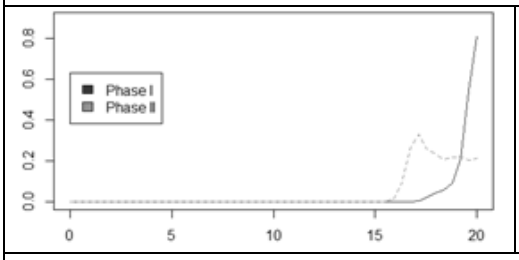
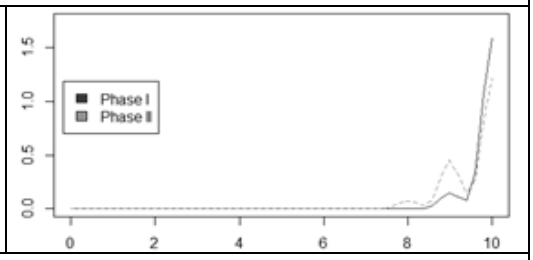
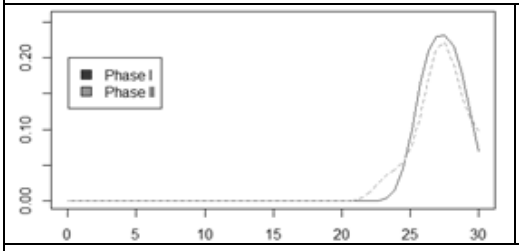
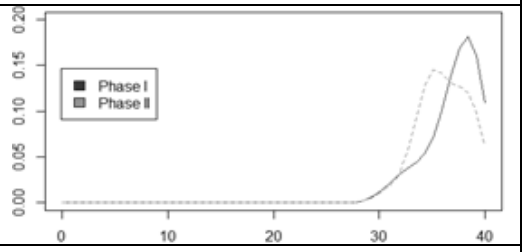
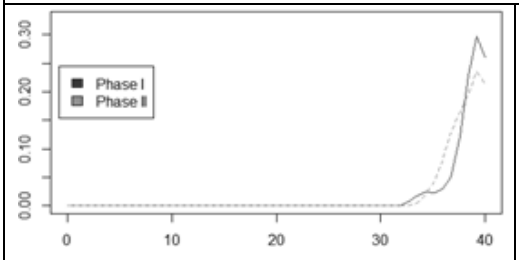
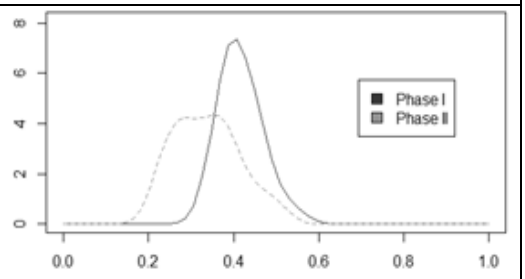
However, it should be noted that the MI technique has helped us identify three additional patients with more severe global impairments among those with missing data who had been previously excluded from the complete cases analysis. This indicated that those individuals had initially been excluded because their missing data was caused by their advanced MS, which, besides motor symptoms, had also included the more advanced cognitive symptoms. We acknowledge that these three individuals with severe MS had been underrepresented in the Chapter Five analyses, however, they were rare cases which explained less than 4% of the total variance, and did not affect the overall results.

Rerunning the Chapter Five analyses on the five MI datasets has shown that the clustering solutions were the most robust for the phase II datasets. Due to the lack of clear numbers of clustering solutions, the clustering patterns for phase I and for longitudinal data were explained better once the large clusters with the majority of patients were contrasted against the many small clusters. The small clusters had included the patients whose cognitive symptoms were more progressive, and most commonly the progression was expressed by worsening verbal memory and information processing speed, and to a lesser extent by worsening visuoperceptual functions. The observed associations between SPMS course and developing deficits in the information processing speed and memory domains have been discussed in more detail in the discussion chapter, Chapter Eight.

Appendix F. Distributions of performance scores at phases I and II

Figure 1. The distribution of controls' performance scores on each item at phases I and II



VOSP2 (N=23)		**VOSP3** (N=23)
		
VOSP6 (N=23)		**VOSP8** (N=23)
		
BORB2 (N=23)		**BORB5** (N=23)
		
TROG (N=21)		**BC.Index** (N=22)
		

Note. The solid line represents distribution of scores at phase I and the dashed line represents distribution of scores at phase II.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

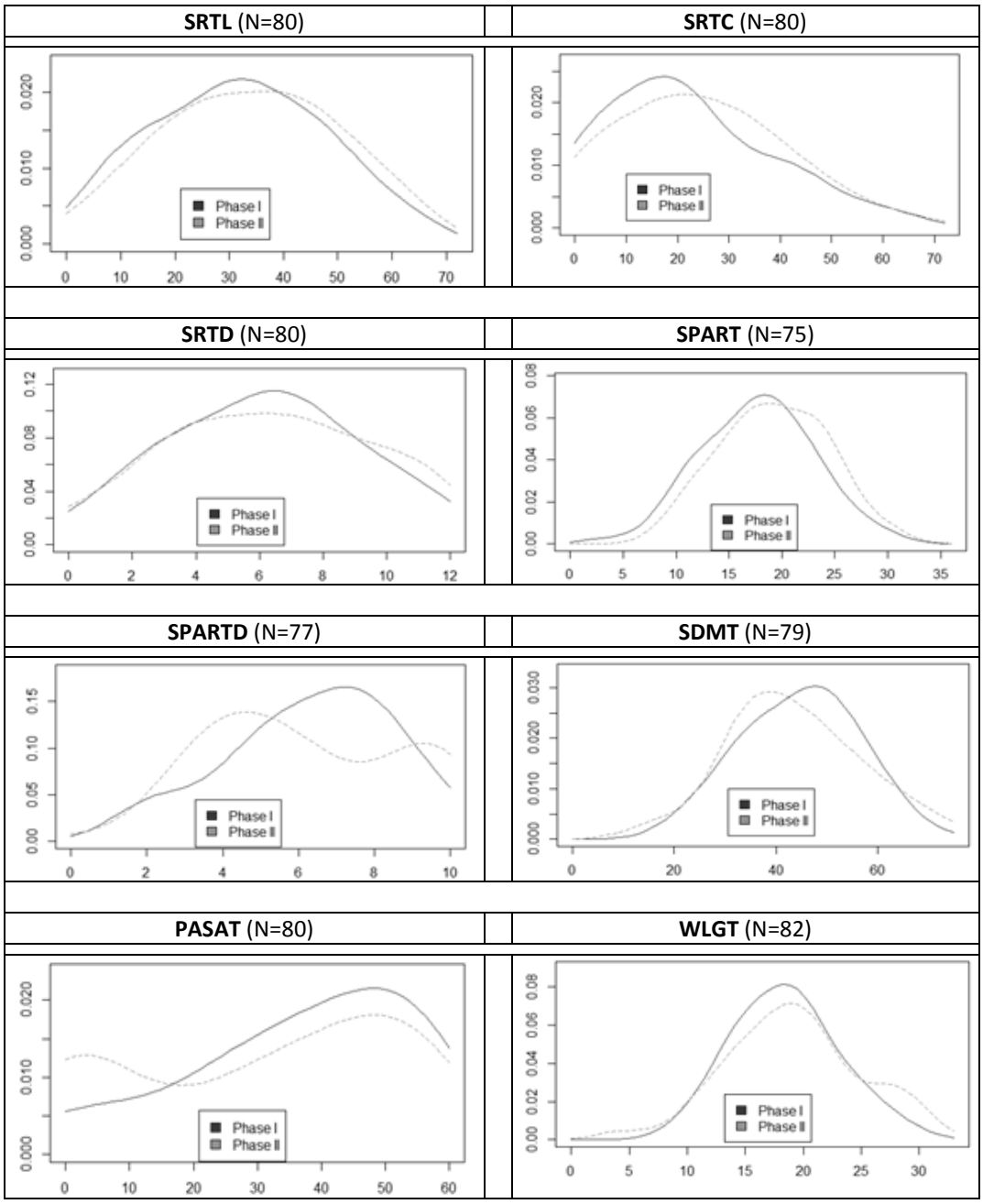
Table 1. Comparisons of distribution of controls' scores on cognitive tests between phase I and phase II.

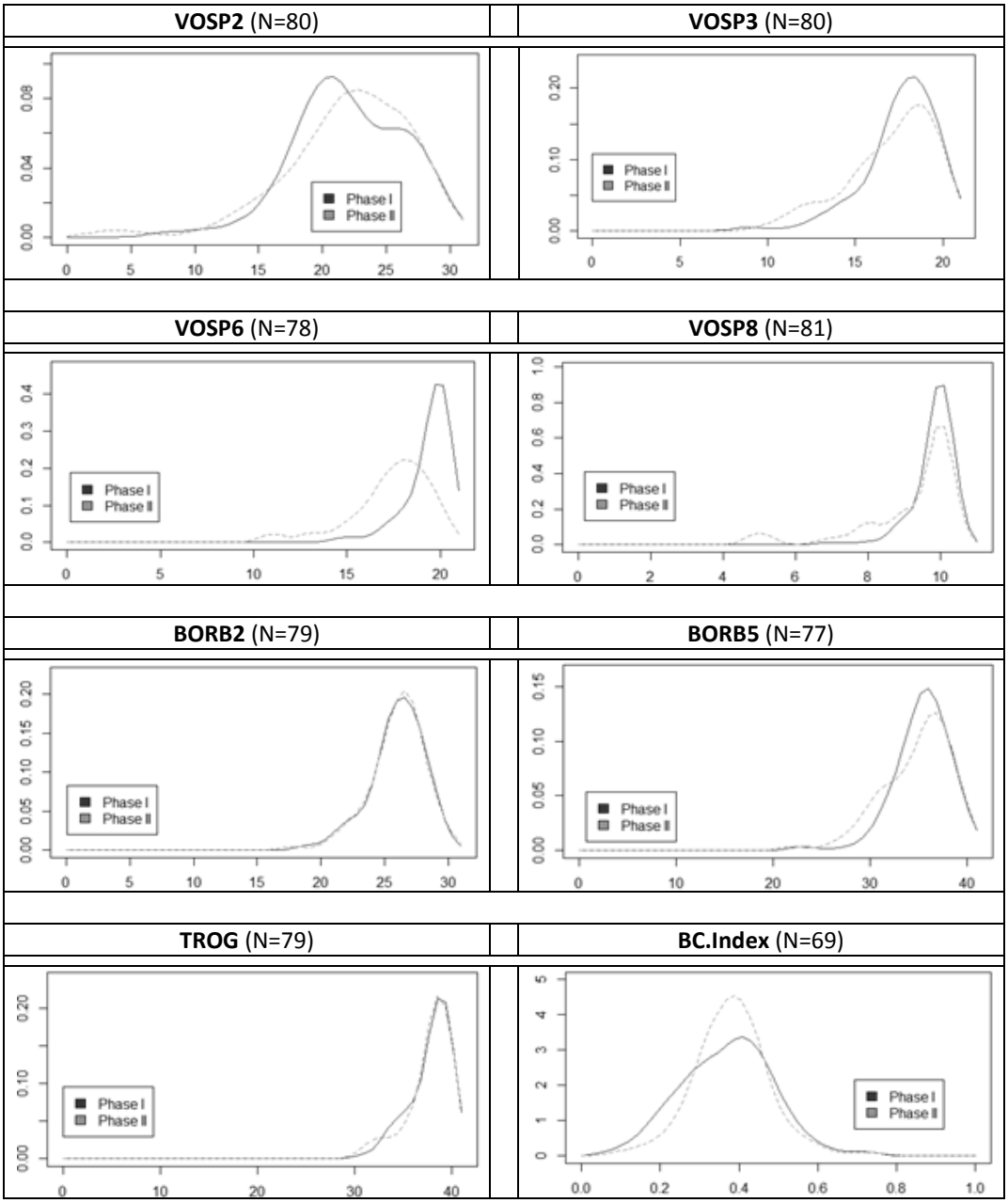
Cognitive test	Difference between score distributions
SRTL (N=23)	D = 0.174, p = 0.877
SRTC (N=23)	D = 0.13, p = 0.99
SRTD (N=23)	D = 0.217, p = 0.649
SPART (N=17)	D = 0.235, p = 0.734
SPARTD (N=23)	D = 0.391, p = 0.059
SDMT (N=22)	D = 0.273, p = 0.386
PASAT (N=22)	D = 0.182, p = 0.86
WLGT (N=23)	D = 0.348, p = 0.124
VOSP2 (N=23)	D = 0.348, p = 0.124
VOSP3 (N=23)	D = 0.087, p = 0.999
VOSP6 (N=23)	D = 0.652, p < 0.001***
VOSP8 (N=23)	D = 0.217, p = 0.649
BORB2 (N=23)	D = 0.13, p = 0.99
BORB5 (N=23)	D = 0.358, p = 0.124
TROG (N=21)	D = 0.19, p = 0.841
BC.Index (N=22)	D = 0.591, p < 0.001 ***

Note. Comparisons were performed using 2-sample Kolmogorov – Smirnov test, 2-sided

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

Figure 2. The distributions of patients' scores on each test item at phases I and II





Note. The solid line represents distribution of scores at phase I and the dashed line represents distribution of scores at phase II.

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

Table 2. Comparisons of distributions of patient scores on cognitive tests between phase I and phase II

Cognitive test	Difference between score distributions
SRTL (N=80)	D = 0.099, p = 0.824
SRTC (N=80)	D = 0.136, p = 0.444
SRTD (N=80)	D = 0.074, p = 0.979
SPART (N=75)	D = 0.16, p = 0.292
SPARTD (N=77)	D = 0.143, p = 0.412
SDMT (N=79)	D = 0.114, p = 0.684
PASAT (N=80)	D = 0.15, p = 0.329
WLGT (N=82)	D = 0.098, p = 0.83
VOSP2 (N=80)	D = 0.137, p = 0.436
VOSP3(N=80)	D = 0.175, p = 0.172
VOSP6 (N=78)	**D = 0.603, p < 0.001 *** **
VOSP8 (N=81)	**D = 0.225, p = 0.035 ***
BORB2 (N=79)	D = 0.025, = 0.999
BORB5 (N=77)	D = 0.117, p = 0.669
TROG (N=79)	D = 0.038, p = 0.999
BC.Index (N=69)	D = 0.174, p = 0.248

Note. The comparisons were performed with 2-sample Kolmogorov – Smirnov test, 2-sided

Abbreviations: SRT – Selective Reminding Test (SRTL – Long Term Storage, SRTC – Consistent Long Term Retrieval, SRTD – Delayed Retrieval), SDMT – Symbol Digits Modalities Test, PASAT – Paced Auditory Serial Addition Test, SPART – 10/36 Spatial Recall Test (SPART– items correct at learning stage, SPARTD– items correct at delayed recall), WLGT – Category Animal Fluency task, VOSP – Visual Object and Space Perception Battery (VOSP2 – Silhouette Naming Task, VOSP3 – Object Decision Task, VOSP6 – Position Discrimination Task, VOSP8 – Cube Counting Task,) BORB – Birmingham Object Recognition Battery (BORB2 – Line Length Matching Task, BORB5 – Position of Gap Matching Task, TROG – Test of Reception of Grammar, BC – Boston Cookie Test (BC.Index – Index of ratio of picture variables to number of words).

Appendix G. Standardized scores of longitudinal change

Table 1. Averaged z-scores of individual patient (n=61) change for each domain

Averaged z-score	Definition
> 3	great improvement
1.5 to 3	moderate improvement
0.5 to 1.5	mild improvement
-0.5 to 0.5	No change
-1.5 to -0.5	mild deterioration
-3 to -1.5	moderate deterioration
< -3	severe deterioration

ID	Verbal memory	Visuospatial memory	Processing speed	Visuoperceptual	Language
1	-0.13	-0.11	-0.38	0.22	0.74
2	0.52	-0.62	-2.11	-0.95	0.74
3	0.24	-1.01	-1.51	-1.7	-0.68
4	0.5	-0.23	-0.35	-0.32	1.68
5	1.17	-0.1	-0.04	-0.38	1.2
6	-1.49	-1.26	0.5	0.38	0.74
7	1.57	2.11	-0.55	-0.59	1.21
8	-0.36	-0.61	-0.6	0.02	2.16
9	-1.71	-2.3	-1.34	0.18	1.2
10	-2.57	-0.48	-2.55	-0.44	0.73
11	0.01	0.81	-0.35	-0.42	1.68
12	-1.07	-0.36	0.09	-0.43	0.26
13	2.16	0.03	-0.64	-0.21	0.26
14	0.77	0.15	-1.06	-0.19	1.66
15	-1.27	0.68	-1.82	0.23	2.61
16	0.13	-1.4	-4.1	-0.58	-0.21
17	0.16	-1.01	-0.65	-0.25	0.73
18	-0.22	0.81	0.4	0.02	1.2
19	-0.38	-0.36	-0.62	0.35	0.72
20	0.47	-1.28	0.46	0.19	0.25
21	0.63	0.81	-0.89	0.03	1.19
22	1.05	-0.23	1.1	0.83	2.61
23	1.63	-2.57	-1.07	0.94	0.73
24	-0.08	-1.4	-1.34	0.63	-0.65
25	0.31	0.04	-0.82	0.78	1.2
26	1.72	1.19	0.25	-0.03	1.68
27	0.3	0.95	0.93	0.52	1.21
28	-0.1	1.07	-1.36	0.14	1.67
29	-0.86	-1.53	-1.73	0.38	-0.19
30	0.13	-2.43	-0.06	0.33	1.68

	Verbal memory	Visuospatial memory	Processing speed	Visuoperceptual	Language
31	-1.37	1.07	0.07	0.51	0.74
32	-0.12	-0.48	-0.17	-0.33	0.74
33	-0.48	-1.01	-0.14	-0.63	1.2
34	0.01	-1.52	-0.32	-0.32	0.73
35	-0.3	0.56	-0.57	-0.67	0.74
36	0.03	-1.79	-0.99	0.02	0.75
37	0.49	0.04	-0.22	0.31	0.74
38	1.52	0.03	0.84	-0.2	0.27
39	2.14	0.03	0.41	1.03	-0.21
40	1.14	1.07	-0.19	-0.6	1.22
41	-0.84	-1.91	0.27	0.16	-2.57
42	0.24	1.08	-1.93	-1.69	0.72
43	-0.37	-2.18	-0.43	-0.18	0.25
44	-0.09	0.42	-0.07	-0.29	1.2
45	-1.65	2.23	-1.31	0.66	-0.67
46	-1.94	-0.36	0.49	-1.59	0.73
47	0.42	-0.23	0.04	-0.52	2.6
48	0.43	-0.1	-1.51	0.82	0.72
49	-0.55	-0.36	-1.02	-1.73	0.25
50	0.32	-1.53	-2.03	-1.47	-0.22
51	1.79	-0.23	-0.64	0.24	0.72
52	-0.52	0.28	-0.71	-0.48	1.2
53	0.39	0.03	0.43	-2.41	1.68
54	-0.72	-1.14	-0.51	-0.83	-1.63
55	0.44	0.69	-0.96	-0.32	-0.22
56	-0.31	-1.66	-0.49	-1.7	2.62
57	2.43	0.42	0.4	-0.02	0.27
58	2.8	1.19	-0.69	0.91	0.71
59	0.51	-1.4	0.41	0.4	0.75
60	2.67	0.55	0.56	0.53	0.74
61	1.59	0.04	0.17	-1.24	0.73

Appendix H. Standardized scores of actual change, and scores of perceived change for each individual

Table 1. Patient (n=61) actual and perceived change for each group of cognitive tests

Averaged z-score	Definition
> 3	great improvement
1.5 to 3	moderate improvement
0.5 to 1.5	mild improvement
-0.5 to 0.5	No change
-1.5 to -0.5	mild deterioration
-3 to -1.5	moderate deterioration
< -3	severe deterioration

	Actual change (z-scores)			Perceived change (%)		
ID	BRBN battery	Visuoperceptual	Language	BRBN battery	Visuoperceptual	Language
1	-0.21	0.22	0.74	-15	0	0
2	-0.74	-0.95	0.74	-50	-15	-5
3	-0.76	-1.7	-0.68	-50	-27	-30
4	-0.03	-0.32	1.68	0	0	-10
5	0.34	-0.38	1.2	-5	-10	-10
6	-0.75	0.38	0.74	-25	0	-5
7	1.04	-0.59	1.21	-5	-5	5
8	-0.52	0.02	2.16	-10	-10	0
9	-1.78	0.18	1.2	-20	0	-5
10	-1.87	-0.44	0.73	-27	-30	-15
11	0.16	-0.42	1.68	-5	-5	-5
12	-0.45	-0.43	0.26	10	10	0
13	0.52	-0.21	0.26	-15	-20	0
14	-0.05	-0.19	1.66	-10	-30	0
15	-0.8	0.23	2.61	-45	15	10
16	-1.79	-0.58	-0.21	-20	-15	0
17	-0.5	-0.25	0.73	0	-10	0
18	0.33	0.02	1.2	-5	-20	0
19	-0.45	0.35	0.72	-25	-5	-10
20	-0.12	0.19	0.25	-10	-10	0
21	0.18	0.03	1.19	-10	-5	0
22	0.64	0.83	2.61	-10	5	0
23	-0.67	0.94	0.73	-10	0	-10
24	-0.94	0.63	-0.65	-20	-5	0
25	-0.16	0.78	1.2	-25	-5	-15

	Actual change			Perceived change		
ID	BRBN battery	Visuoperceptual	Language	BRBN battery	Visuoperceptual	Language
26	1.05	-0.03	1.68	0	-10	0
27	0.73	0.52	1.21	20	0	0
28	-0.13	0.14	1.67	-25	-30	-40
29	-1.37	0.38	-0.19	-10	0	0
30	-0.79	0.33	1.68	-10	0	0
31	-0.08	0.51	0.74	-5	-10	0
32	-0.26	-0.33	0.74	-8	0	0
33	-0.54	-0.63	1.2	0	0	0
34	-0.61	-0.32	0.73	-15	0	0
35	-0.1	-0.67	0.74	0	15	-5
36	-0.92	0.02	0.75	0	-5	0
37	0.1	0.31	0.74	-20	0	0
38	0.8	-0.2	0.27	-20	0	-10
39	0.86	1.03	-0.21	-8	-15	0
40	0.67	-0.6	1.22	-10	-5	5
41	-0.83	0.16	-2.57	-30	-25	-10
42	-0.2	-1.69	0.72	-40	-20	30
43	-0.99	-0.18	0.25	-20	10	0
44	0.09	-0.29	1.2	-5	0	0
45	-0.24	0.66	-0.67	-10	-25	-5
46	-0.6	-1.59	0.73	-5	0	0
47	0.08	-0.52	2.6	0	-10	-10
48	-0.39	0.82	0.72	-5	-5	0
49	-0.64	-1.73	0.25	-20	0	-10
50	-1.08	-1.47	-0.22	-35	-9	-7
51	0.31	0.24	0.72	-5	0	0
52	-0.32	-0.48	1.2	-35	0	25
53	0.28	-2.41	1.68	-30	-15	0
54	-0.79	-0.83	-1.63	0	-10	0
55	0.06	-0.32	-0.22	0	0	0
56	-0.82	-1.7	2.62	-20	-15	0
57	1.08	-0.02	0.27	0	0	0
58	1.1	0.91	0.71	0	-10	0
59	-0.16	0.4	0.75	-10	0	0
60	1.26	0.53	0.74	10	-20	0
61	0.6	-1.24	0.73	-25	-15	-25

Note. The actual change is presented in standardized scores. Positive values of perceived change indicate perceived improvement, negative values indicate perceived deterioration, and values around 0 indicate perceived stability of cognitive function.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological tests.

Table 2. Control (n=15) actual and perceived change for each group of cognitive tests

Averaged z-score	Definition
> 3	great improvement
1.5 to 3	moderate improvement
0.5 to 1.5	mild improvement
-0.5 to 0.5	No change
-1.5 to -0.5	mild deterioration
-3 to -1.5	moderate deterioration
< -3	severe deterioration

	Actual change			Perceived change		
Case	BRBN battery	Visuoperceptual	Language	BRBN battery	Visuoperceptual	Language
1	0.29	-0.37	-0.02	10	0	7
2	0.32	-0.68	-0.56	5	0	0
3	0.55	-0.08	-0.46	-50	-10	0
4	0.3	0.02	-0.53	0	0	0
5	0.11	-1.07	-0.61	-20	0	-10
6	-0.22	-0.33	-0.23	-25	5	-5
7	0.13	0.21	1.14	-30	-10	-10
8	-0.26	1.02	-0.14	-10	0	0
9	0.51	0.11	1.15	-38	0	7
10	-0.12	-0.41	0.2	-28	10	10
11	0.16	0.19	-1.59	0	0	5
12	-0.21	-0.34	0.83	0	0	-5
13	0.3	0.44	1.06	7	1	2
14	0.06	0.09	-0.48	0	0	0
15	-0.64	-0.25	0.76	-27	-27	-7

Note. The actual change is presented in standardized scores. Positive values of perceived change indicate perceived improvement, negative values indicate perceived deterioration, and values around 0 indicate perceived stability of cognitive function.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological tests.

Appendix I. Awareness of cognitive deficits in pwMS

1. Section Aim

The aim of the following insert was to rerun the Chapter Seven awareness analyses (section 7.3), but this time in search for cross-sectional differences in awareness of cognitive difficulties on the data *at phase II*, instead of focusing on self-perception of *longitudinal change*.

This supplementary analysis investigated the relationships between actual cognitive performance and the patients' own perception of how they performed on the cognitive tests at phase II. The separate but linked analysis revolved around investigation of the factors that may help explain the discrepancies between the actual and perceived cognitive deficits. The supplementary analyses performed in this insert were designed to yield a better understanding of whether the patients had been accurate in estimating their performance, and whether they had been capable to make assumptions about the amount of perceived longitudinal change.

2. Methods

2.1. Participants

For all cross-sectional analyses in this section only the participants with full data on all cognitive assessments at phase II were chosen to employ, the reason behind it was that the self-evaluation data had been collected only at phase II. By restricting the analyses to participants with full data at phase II it was controlled that all participants had gone through the exactly same tests and therefore provided their self-estimates of cognitive performance for the same assessments.

Therefore for all of the cross-sectional analyses of the actual and the perceived performance, the scores of 63 patients and 23 controls were employed. For further analyses to investigate the predictors of self-evaluations of cognitive performance, only the data from the patient participants were used.

2.2. Data format

2.2.1. Actual performance at phase II

The scores of actual performance at phase II were all converted to percentages. For the tests where the maximum value was unknown (SDMT, WLGT and BC.Index), the highest value from the control and patient samples was used to create the 100% score. The percentage values were calculated separately by averaging the scores of the component tests for each of these groups of tests: the BRBN battery, visuoperceptual and language tests.

2.2.2. Perceived performance

The estimates of perceived performance at phase II were collected on three groups of cognitive tests: the BRBN battery, and the visuoperceptual and language tests. Higher self-reported values indicated better perceived performance and the range of estimations at phase II was on a scale from 0 to 100.

2.3. Group differences in performance at phase II

The scores of actual performance at phase II were converted to percentages and were then compared between the patient and control participants with the aim of investigating any differences that could be found *between the patients and controls* in how well they had performed at phase II on the three groups of cognitive tests. All comparisons were performed with the Mann-Whitney U-test. The dependent variables were the percentage scores of actual performance on the BRBN tests, the visuoperceptual tests and the language tests; and the independent variable was being a patient or a control.

As a next step the actual performance at phase II on the three groups of tests was also compared for the patient and control samples separately. The aim of this analysis was to investigate whether there had been differences in actual performance *between the three groups of tests*. Namely, this way it was tested whether the participants had performed on some groups of tests better than on others. Those analyses were performed with the Related-Samples Friedman's analysis of variance by ranks, first for the patient participants, then separately for the control participants.

2.4. Group differences in perceived performance at phase II

After analyzing the actual performance, the next step was to investigate whether there have been differences between the self-perceptions of performance at phase II. The analyses of perceived performance were conducted in the same manner as the analyses of actual performance, but this time the percentage of perceived rather than actual performance was analysed.

First the self-estimates of performance at phase II *between the patient and control participants* were compared. It was done to investigate whether there wed been any differences between the patients and the controls in how well they thought that they had performed at phase II. The performance of the patients and the controls on all three groups of cognitive tests was compared using the Mann-Whitney U-test. The dependent variables were the scores of perceived performance at phase II on the BRBN tests, visuoperceptual and language tests; and the independent variable was being a patient or a control.

The perceived performance on the three groups of cognitive tests at phase II was also compared for the patient and control samples separately. This was done in order to investigate whether there had been any differences in the perceived performance *between the three groups of cognitive tests* at phase II. These analyses were performed with the Related-Samples Friedman's analysis of variance by ranks twice, first for the patient participants, and then separately for the controls.

2.5. Relationship between the actual and the perceived performance at phase II

Having analysed the actual performance and the perceived performance at phase II separately, the next step was to investigate the relationship between them. These analyses aimed to answer the question whether the participants were correct in estimating their own performance. This was done separately for each group of cognitive tests in both patients and controls.

In the first set of analyses it was investigated how correct the participants were in estimating their performance. This was done by comparing the estimated performance to their actual performance using the Related Samples Wilcoxon Signed Rank Test.

In the second set of analyses we aimed to investigate the extent to which the actual and the perceived performance have been associated. It was chosen to run multiple correlations between the measures of actual and perceived performance. Spearman correlation was employed due to non-parametric distributions of both the actual and perceived scores in the patient and control cohorts.

2.6. Predictors of perceived performance

The next step was to investigate the factors considered to be related to self-perception of cognitive performance at phase II in the patient population. The following multiple linear regression model for the patient participants was devised:

$$\textit{Self-evaluation of cognitive performance} = \alpha + \beta_1 (\textit{depression score}) + \beta_2 (\textit{neurological disability score}) + \beta_3 (\textit{MS impact score})$$

The linear regression model predicted that lower self-evaluations would be associated with higher levels of depression, neurological disability and MS impact. It was anticipated that these models would have similar predictive values on all groups of cognitive tests.

In this model the depression was measured with the BDI-II scale, the neurological disability was measured with the EDSS scale, and the MS impact was measured with the MSIS-29 scale. All data was collected at phase II.

3. Results

In this chapter the percentage scores on the individual BRBN, visuoperceptual and language tests were averaged to produce the average values of performance for each group of tests for both patient and control participants.

Table 1. Descriptives of actual and perceived performance on the groups of cognitive tests for patient and control participants

	Patients (n=63)		Controls (n=23)	
	Actual performance	**Perceived performance**	**Actual performance**	**Perceived performance**
BRBN	57% (15.73)	45% (22.71)	73% (10.51)	58% (19.41)
Visuoperceptual	87% (5.67)	62% (17.78)	90% (5.43)	72% (12.94)
Language	74% (17.78)	72% (16.56)	72% (5.34)	79% (10.85)

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological tests

As it can be seen in Table 1, there had been few inaccuracies in estimating the performance at phase II. Overall there had been a tendency to underestimate the cognitive performance for both groups of participants.

3.1. Group differences in actual performance at phase II

First the actual performance was compared between the patient and control participants on the three groups of cognitive tests. This was done to investigate whether there had been differences between the patients and controls in how they've performed on the tests at phase II.

As it can be seen from Figure 1, the controls had performed statistically significantly better than the patients on the BRBN and visuoperceptual tests at phase II. However, there had been no differences between the groups on the language tests. On average, the patients and the controls had performed the same on the language tests at phase II, but there was a wider range of scores in the patient group, indicative of presence of both the patients who performed very well and those who performed poorly.

Figure 1. Comparison of performance percentages between control and patient participants for each group of cognitive tests at phase II

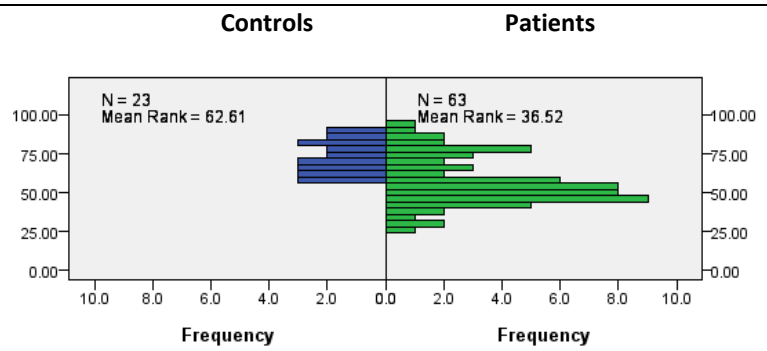
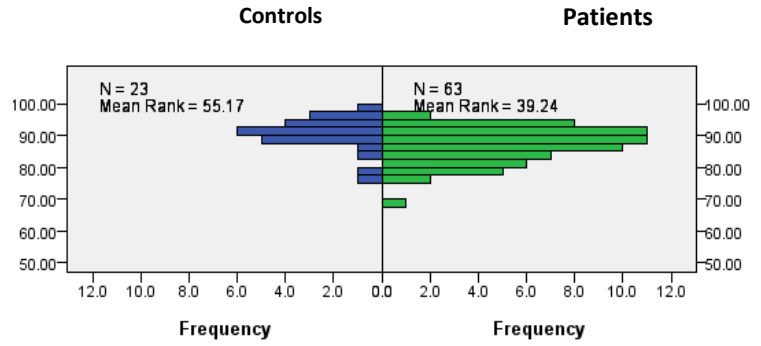
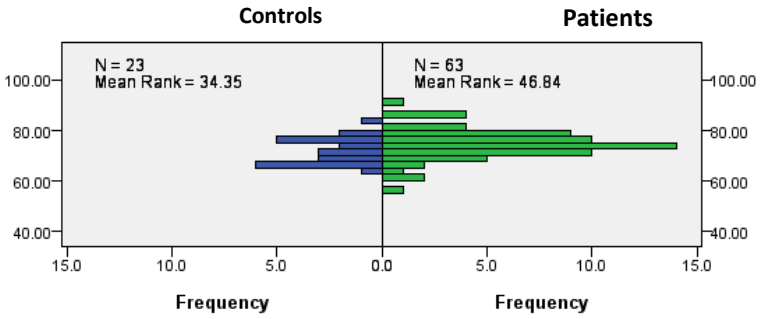
Group of tests		Difference
BRBN battery		**U = 285.00, p < 0.001**
Visuo-perceptual		**U = 456.00, p = 0.009**
Language		U = 935.00, p = 0.060

Figure 1 shows that the controls have performed statistically significantly better on BRBN and visuoperceptual tests than patients. There was no significant difference in the performance on the language tests between the groups at phase II.

Note. Group comparisons were performed using the Mann-Whitney U-test

Abbreviations: BRBN – Brief Repeatable Battery of Neuropsychological tests

3.2. Group differences in perceived performance at phase II

As the next step the self-evaluations of performance on the tests at phase II were compared between the groups. This was done in order to investigate whether there had been any differences in how well the patients and controls had thought that they had performed at phase II.

As it can be seen from Figure 2, on average the controls thought that they had performed better than the patients, and this difference was statistically significant only on BRBN and visuoperceptual tests. The same tendency was observed on the language tests, even though the group differences did not reach statistical significance.

Figure 2. Comparison of self-estimates of perceived performance between control and patient participants for each group of cognitive tests at phase II

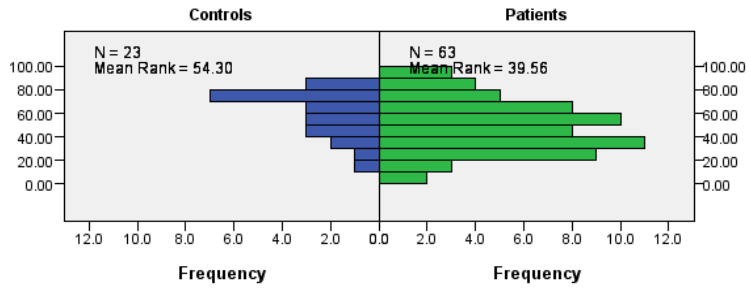
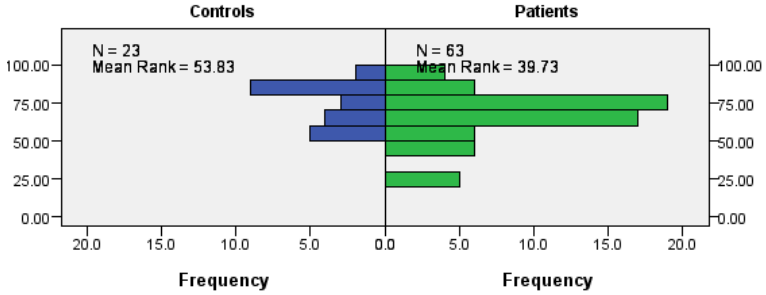
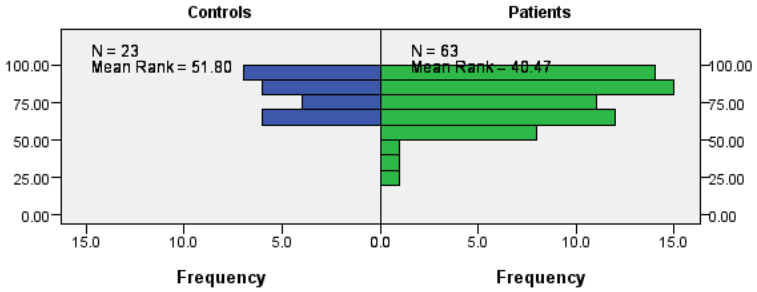
Group of tests		Difference
BRBN battery		**U = 476.00, p = 0.015**
Visuo-perceptual		**U = 487.00, p = 0.020**
Language		U = 533.00, p = 0.060

Figure 2 shows that on average the controls thought that they had performed better than the patients, and this difference was statistically significant only on BRBN and visuoperceptual, but not the language tests.

Note. Group comparisons were performed using the Mann-Whitney U-test

Abbreviations: BRBN – Brief Repeatable Battery of Neuropsychological tests

3.3. Relationship between the actual and the perceived performance at phase II

Having analysed the actual performance and the perceived performance at phase II separately, the next step was to investigate the relationship between them. These analyses aimed to answer the question whether the participants were correct in estimating their own performance.

3.3.1. Perceived and actual performance in the control sample

As it can be seen from Figure 3, the controls had a tendency to evaluate their performance worse than it actually was, and this difference was statistically significant on the BRBN and visuoperceptual tests. This means that overall the controls had thought that they performed the tasks around 58% and 72% correct, when in reality they were on average 73% and 90% correct (Table 1).

However, on the language tests the difference took an opposite direction as the controls had thought that they had performed better than they actually had; although this discrepancy was less often observed than those on BRBN and visuoperceptual tests.

Overall the controls had a tendency to underestimate their performance on the BRBN and visuoperceptual tests, but to overestimate on the language tests.

As it can be seen from Figure 4, the performance scores were linked to self-evaluation scores only on the BRBN battery tests. This was indicative that the estimations of performance on visuoperceptual and language tests were not related to the actual performance in the control sample. It could be argued that the self-evaluations on those tests could be explained better by other variables.

Figure 3. Comparisons between actual and perceived performance at phase II for control participants

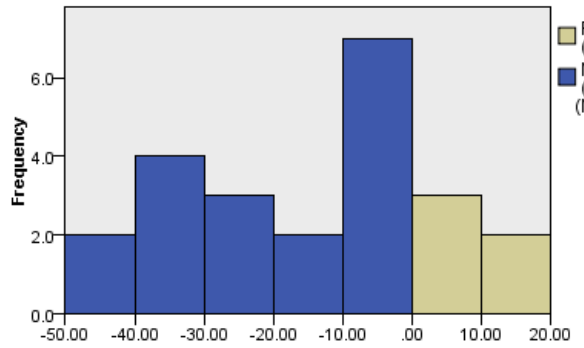
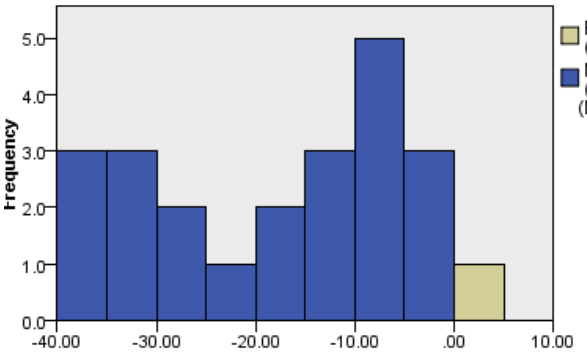
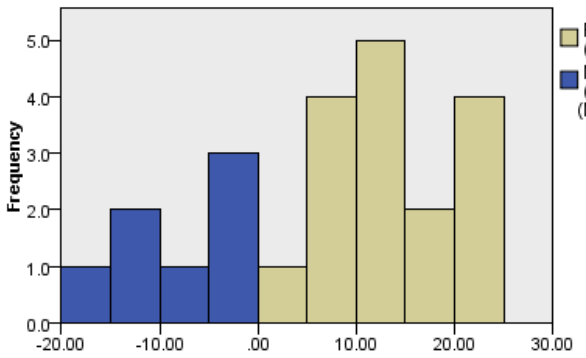
Cognitive tests	Controls (n=23)	Difference
BRBN battery		**Z = 33.00, p = 0.001**
Visuo-perceptual		**Z = 2.00, p < 0.001**
Language		**Z = 218.00, p = 0.015**

Figure 3 shows that the controls had a tendency to underestimate their performance on the BRBN and visuoperceptual tests, but to overestimate on the language tests

Note. The comparisons were performed using the Related Samples Wilcoxon Signed Rank Test.

Abbreviations: BRBN – Brief Repeatable Battery of Neuropsychological Tests

Figure 4. Relationships between actual and perceived performance on the three groups of cognitive tests at phase II

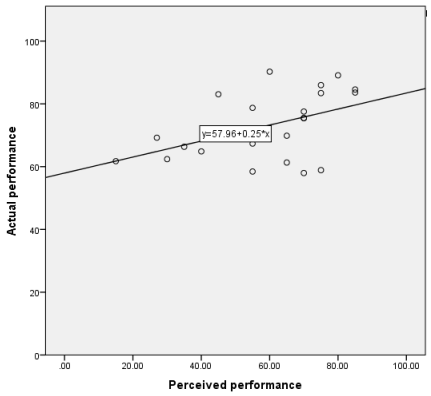
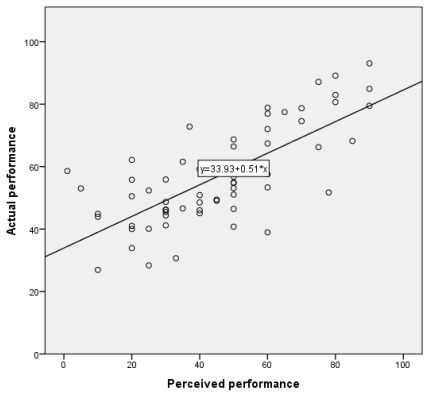
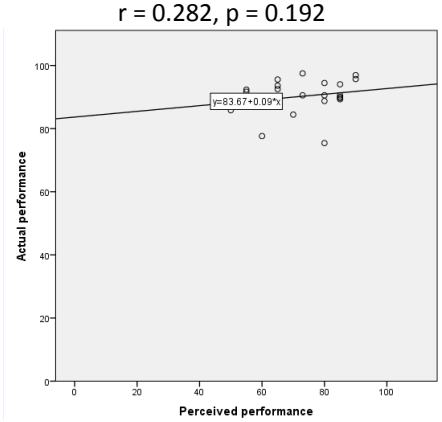
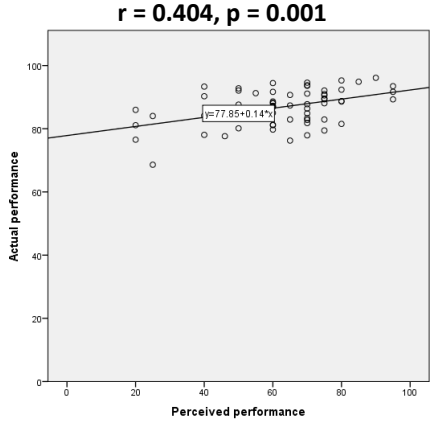
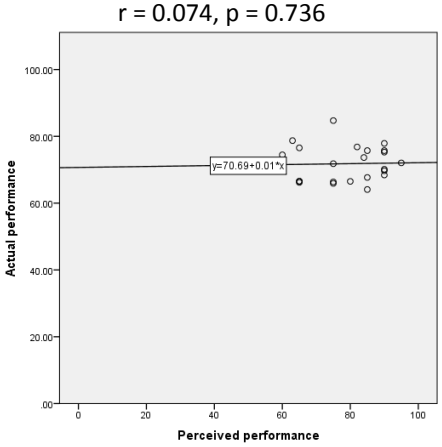
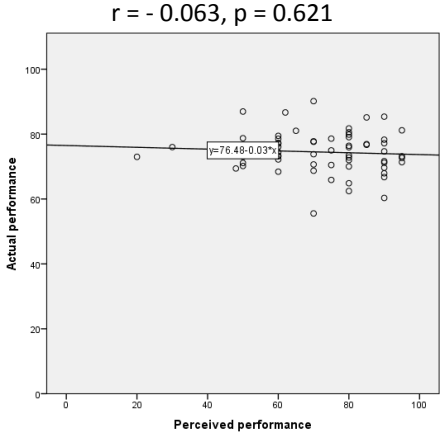
Tests	Controls (n=23)	Patients (n=63)
BRBN battery	**r = 0.471, p = 0.023** 	**r = 0.689, p < 0.001** 
Visuo-perceptual	r = 0.282, p = 0.192 	**r = 0.404, p = 0.001** 
Language	r = 0.074, p = 0.736 	r = - 0.063, p = 0.621 

Figure 4 shows that the performance scores were linked to self-evaluation scores only on the BRBN battery for both groups of participants, and only for patients on the visuoperceptual tests.
Note. The relationships were examined using the Spearman correlation coefficient
Abbreviations: BRBN – Brief Repeatable Battery of Neuropsychological Tests

3.3.2. Perceived and actual performance in the patient sample

As it can be seen from Figure 5, the patients had exhibited a non-different pattern from the controls, as they also had a tendency to evaluate their performance worse than it actually was. This difference was statistically significant on the BRBN and visuoperceptual tests.

However, on the language tests the patients had underestimated and overestimate their performance an equal amount of times, and with a similar degree of inaccuracies.

As it can be seen from Figure 4, the performance scores were linked to self-evaluation scores only on the BRBN and visuoperceptual tests. This was indicative that the estimations of performance on language tests were not related to the actual performance in the patient sample. It could be argued that the self-evaluations on those tests could be explained better by other variables.

Figure 5. Comparisons between actual and perceived performance at phase II for patient participants

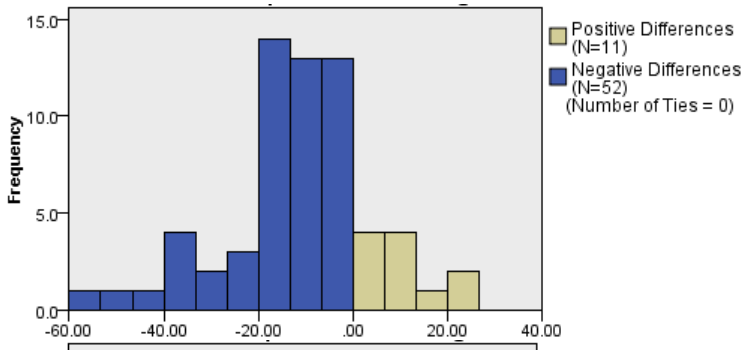
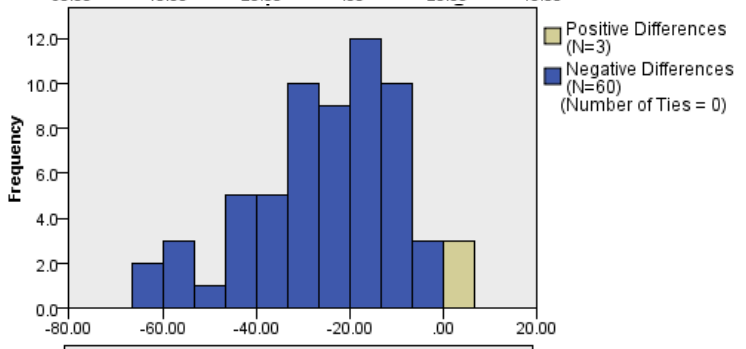
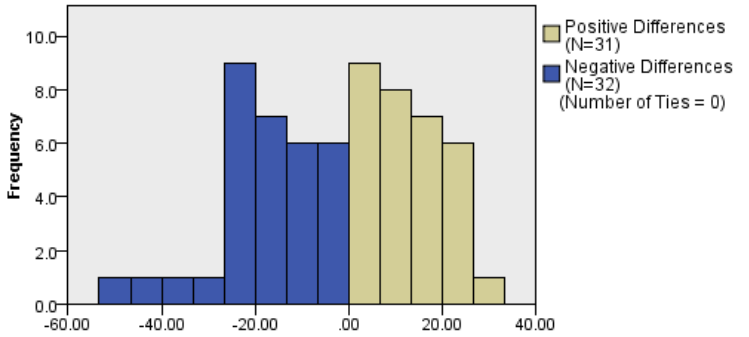
Cognitive tests	Patients (n=63)	Difference
BRBN battery		**Z = 273.00, p < 0.001**
Visuo-perceptual		**Z = 9.50, p < 0.001**
Language		Z = 875.00, p = 0.363

Figure 5 shows that the patients had a tendency to underestimate their performance on the BRBN and visuoperceptual tests.

Note. The comparisons were performed using the Related Samples Wilcoxon Signed Rank Test.

Abbreviations: BRBN – Brief Repeatable Battery of Neuropsychological Tests

3.4. Predictors of perceived performance

As it can be seen from previous analyses earlier in this appendix, the patient and control participants tend to think that they have performed worse than they actually have on the BRBN and visuoperceptual, but not the language tests. Moreover, the actual performance and perceived performance were only correlated on the BRBN and visuosperceptual tests for the patient participants; and the perceived performance could not be fully explained by the actual performance only.

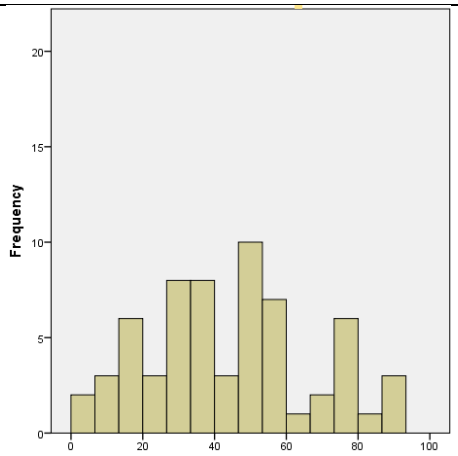
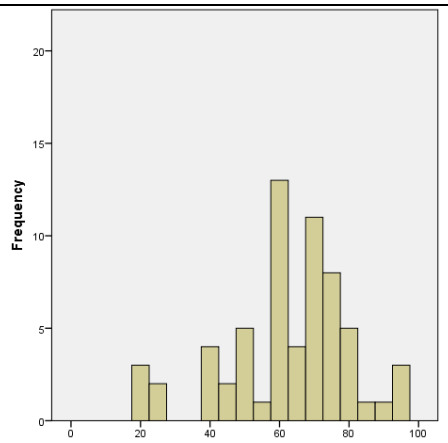
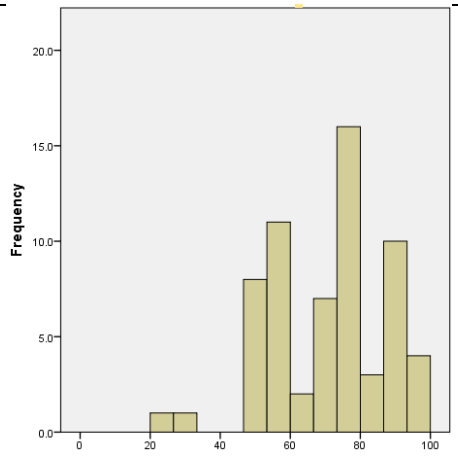
The next step was to investigate the factors considered to be related to self-perception of cognitive performance at phase II in the patient population. The following analyses were devised in order to explain what other factors beside the actual performance influence the patients' decision about whether they have performed well or poorly.

This was examined by running a multiple linear regression model for the patient participants with the perceived performance as the criterion variable and measurements of depressive symptomology, neurological disability and MS impact as predictor variables.

3.4.1. Analysis of univariate normality of criterion variables

The first step before running multiple linear regression analyses was to check for univariate normality of the criterion variables. As it can be seen in Figure 6, the Shapiro-Wilk statistics indicated that most of the criterion variables haven't been normally distributed. However, upon visual inspection it was decided that the distributions were sufficiently normal, as all of them were unimodal with no extreme skew. Therefore it was decided to proceed with running the linear regression models.

Figure 6. Distribution of criterion variables for linear regression models of perceived performance at phase II for patient participants

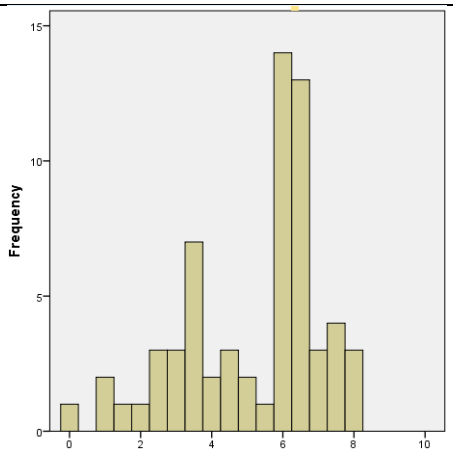
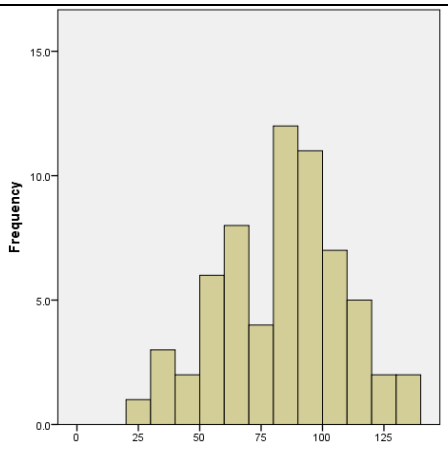
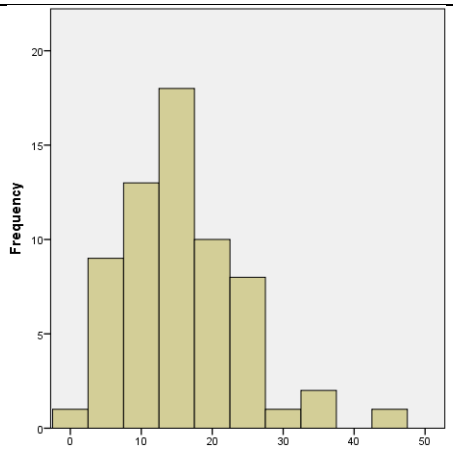
Cognitive tests	Mean (SD)	Shapiro-Wilk test	Distribution
BRBN battery	45.46 (22.71)	W = 0.971, p = 0.148	
Visuoperceptual	62.48 (17.78)	**W = 0.942, p = 0.005**	
Language	71.67 (16.56)	**W = 0.933, p = 0.002**	

Abbreviations: BRBN – Brief Repeatable Battery of Neuropsychological Tests

3.3.2. Analysis of univariate normality of predictor variables

As it can be seen from the distributions of the predictor variables in Figure 7, the Shapiro-Wilk statistics indicated that the EDSS and BDI-II variables haven't been normally distributed. It was evident that a large number of patients had the EDSS scores of 6 and 6.5. However, upon visual inspection it was decided that the distributions were sufficiently normal, as all of them were unimodal with a sufficient range of scores. Therefore it was decided to proceed with including them into the linear regression models.

Figure 7. Distribution of predictor variables for linear regression models of perceived performance at phase II for patient participants

Variable	Mean (SD)	Shapiro-Wilk test	Distribution
EDSS	5.19 (1.94)	**W=0.914, p < 0.001**	
MSIS-29	82.87 (24.91)	W = 0.984, p = 0.602	
BDI-II	15.30 (8.87)	**W = 0.950, p = 0.012**	

Abbreviations: EDSS – Expanded Disability Status Scale; MSIS-29 – 29-item Multiple Sclerosis Impact Scale; BDI-II – Beck's Depression Inventory 2nd edition.

3.3.3. Predictors of perceived performance

Prior to running the linear regression models it was necessary to identify the predictor variables worthy of inclusion, as due to small sample size (63 patients) and insufficient power it was not recommended to include more than two predictor variables for patient models.

This was done by examining relationships between the criterion and predictor variables, and only those predictor variables that were shown to be related to the criterion variables were included in the multiple regression models.

As it can be seen from Table 2, in our patient sample the perceived performance was mainly linked to MS impact. Neurological disability was only linked to perceived performance on the visuoperceptual and language tests, and levels of depression was not likely to have influence on the self-evaluations of perceived performance.

Table 2. Correlations between predictor and criterion variables in models of perceived performance for patient participants at phase II

Predictor variables	BRBN battery	Visuoperceptual	Language
EDSS	r = -0.193, p = 0.130	**r = - 0.300, p = 0.017**	**r = - 0.298, p = 0.018**
MSIS-29	**r = - 0.280, p = 0.026**	**r = - 0.283, p = 0.025**	**r = - 0.270, p = 0.033**
BDI-II	r = 0.008, p = 0.953	r = - 0.026, p = 0.841	r = 0.173, p = 0.176

Table 2 shows that perceived performance was mainly linked to MS impact. Neurological disability was only linked to perceived performance on the visuoperceptual and language tests, and depressive symptomology was not linked to self-evaluations of performance on any domain.

Note. The relationships were examined employing the Spearman correlation

Abbreviations: EDSS – Expanded Disability Status Scale; MSIS-29 – 29-item Multiple Sclerosis Impact Scale; BDI-II – Beck's Depression Inventory 2nd edition.

Based on the findings presented in Table 2 and Figures 6 and 7, it was decided to proceed with running multiple linear regression models with neurological disability and MS impact for perceived performance scores on visuoperceptual and language tests, and a simple linear regression model with MS impact as a sole predictor of perceived performance on the BRBN battery.

I. Predictors of perceived performance on the BRBN battery tests

When perceived performance on BRBN tests was predicted it was found that MS impact (Beta = -0.320, p = 0.011) was a significant predictor. The overall model fit was $R^2 = 0.102$ (Figure 8). This indicated that patients with higher MSIS-29 scores thought that they had performed worse on the BRBN tests at phase II.

Figure 8. The effect of MS impact on the perceived performance on the BRBN tests at phase II for patient participants

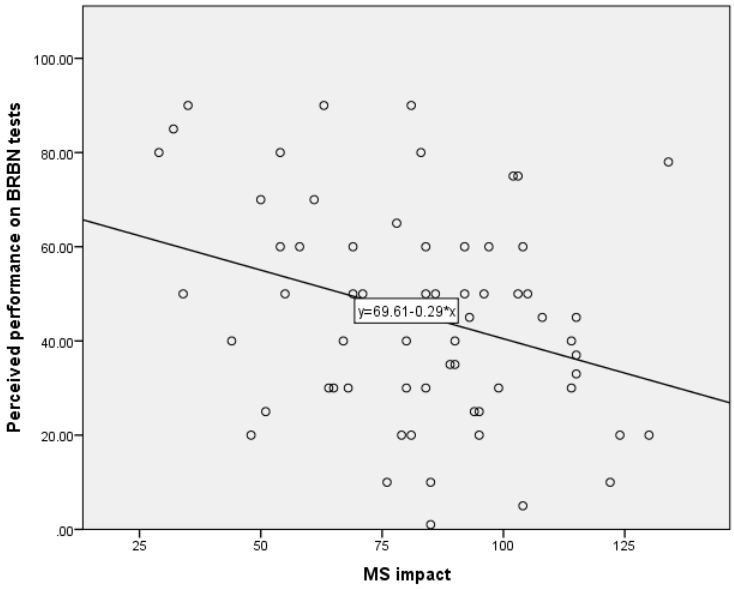


Figure 8 shows that MS impact was predictive of perceived performance on BRBN tests at phase II.

Abbreviations: BRBN – Brief-Repeatable Battery of Neuropsychological tests

II. Predictors of perceived performance on the visuoperceptual tests

Based on the presented associations (Table 2) it was decided to run a linear regression model with perceived performance on visuoperceptual tests as the criterion variable and neurological disability (EDSS score) and MS impact (MSIS-29 score) as predictor variables.

Figure 9. The effects of neurological disability and MS impact on the perceived performance on the visuoperceptual tests at phase II for patient participants

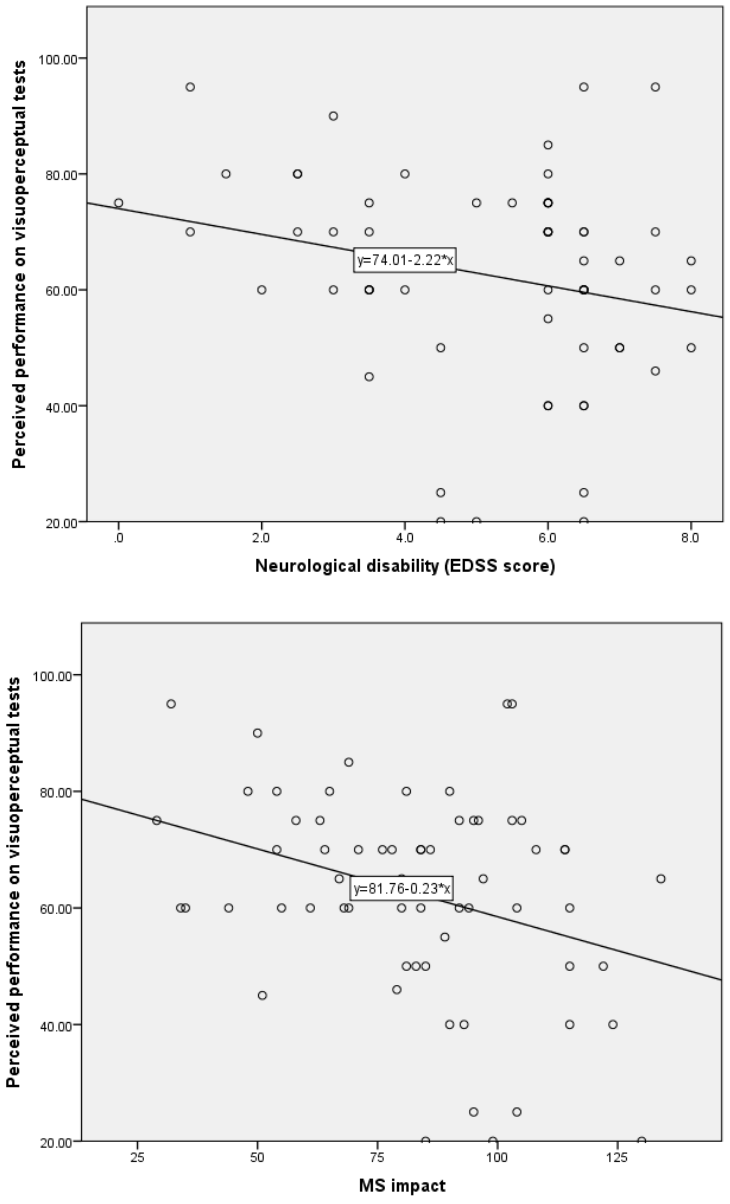


Figure 9 shows that MS impact had more impact on perceived performance than neurological disability on the visuoperceptual tests at phase II

Abbreviations: EDSS – Expanded Disability Status Scale

When perceived performance on visuoperceptual tests was predicted it was found that neither neurological disability (Beta = -0.039, p = 0.815) nor MS impact (Beta = -0.214, p = 0.076) were significant predictors (Figure 9), even though the model itself had significant predictive power (R^2_{adj} = 0.102, p = 0.033).

Since MS impact was the stronger predictor in this model, the model was rerun after excluding neurological disability, with MS impact as a sole predictor variable. In this model when perceived performance on visuoperceptual tests was predicted, it was found that the predictive value of the sole MS impact variable had slightly increased (Beta = -0.233, p = 0.009), and the model fit slightly improved (R^2 = 0.106).

This implied that patients with higher MS impact thought that they had poorer performance on visuoperceptual tests at phase II. Neurological disability had questionable predictive value.

III. Predictors of perceived performance on the language tests

When perceived performance on the language tests at phase II was predicted for patient participants, neither neurological disability (Beta = -0.205, p = 0.221) nor MS impact (Beta = -0.151, p = 0.0367) were significant predictors (Figure 10), even though the model itself had significant predictive power (R^2_{adj} = 0.107, p = 0.034). As both predictor variables had similar coefficients, it was decided not to further investigate their predictive power individually. This implied that patients with higher neurological disability and MS impact were only slightly more likely to give poorer estimates of their performance on the language tasks.

Figure 10. The effects of neurological disability and MS impact on the perceived performance on the language tests at phase II for patient participants

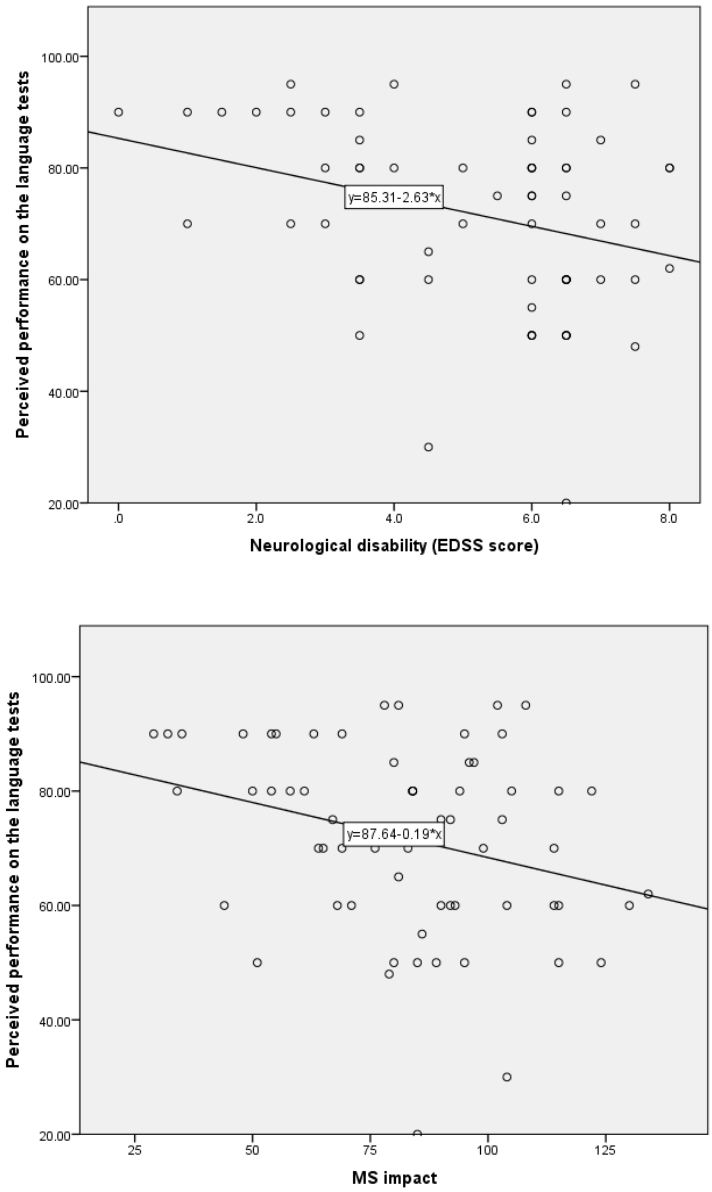


Figure 10 shows that both higher neurological disability and MS impact were to some extent linked to lower perceived performance on the language tests

Abbreviations: EDSS – Expanded Disability Status Scale

4. Section summary

In this section the awareness analyses of Chapter Seven had been rerun on the phase II data investigating the relationships between self-estimations and the actual cognitive performance, as well as the factors that influence self-estimations of performance on different groups of cognitive tests. It was found that overall the patients were able to report their level of cognitive difficulties with the same accuracy as the controls, which allows us to conclude, that the self-estimations of longitudinal change analysed in Chapter Seven had been reliable measures of perceived changes in performance.

The results presented in this insert have shown that overall there had been a tendency to underestimate the cognitive performance for both groups of participants on the BRBN and visuoperceptual tests. On the language tests, however, the controls had slightly more often overestimated their performance, and the patients had both underestimated and overestimated their performance an equal amount of times, and with a similar degree of inaccuracies. On average, the patients and the controls had performed similarly on the language tests, but there has been a wider range of scores in the patient group, indicative of presence of both the patients who performed very well and those who performed poorly, while the performance of the controls was more uniform. This difference in performance could explain different profiles of self-estimations of performance on the language tests between the patient and control participants.

In the subsequent analysis it was found that the performance scores were linked to self-evaluation scores only on the BRBN battery tests for the controls; and on the BRBN battery and visuoperceptual tests for the patients. This was indicative that the estimations of performance on language tests were not related to the actual performance neither in the patient nor in the control sample. It could be argued that the self-evaluations on those tests could be explained better by other variables, rather than the actual performance.

Upon further investigation of the variables linked to self-estimations of cognitive performance at phase II it was found that higher MS impact was linked to poorer perceived performance on all groups of cognitive tests. Besides MS impact, higher neurological disability was associated with poorer perceived performance on the visuoperceptual and language tests, but only in correlational but not regression analyses. These results could be used to conclude that cognitive impairments tend to affect

pwMS and be associated with higher MS impact, while the perception of visuoperceptual and language impairments could be in part linked to neurological disability, such as damage to the optic nerve or dysarthria.

References

Adelman, G., Rane, S. G., & Villa, K. F. (2013). The cost burden of multiple sclerosis in the United States: a systematic review of the literature. *Journal of Medical Economics, 16*(5), 639-647.

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster Analysis. Sage University Paper Series On Quantitative Applications in the Social Sciences 07-044.

Altinkaya, A., Guclu, I., Kurt, E., Bingol, A., Soysal, A., Yandim-Kuscu, D., & Topcular, B. (2012). Long-Term Effects of Cognitive Rehabilitation in Multiple Sclerosis. *Multiple Sclerosis, 18*, 178-179.

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

Amato, M. P., Portaccio, E., Goretti, B., Zipoli, V., Iudice, A., Della Pina, D., . . . Falcini, M. (2010). Relevance of cognitive deterioration in early relapsing-remitting MS: a 3-year follow-up study. *Multiple Sclerosis, 16*(12), 1474-1482.

Amato, M. P., Razzolini, L., Goretti, B., Stromillo, M. L., Rossi, F., Giorgio, A., . . . Portaccio, E. (2013). Cognitive reserve and cortical atrophy in multiple sclerosis A longitudinal study. *Neurology, 80*(19), 1728-1733.

Archibald, C. J., & Fisk, J. D. (2000). Information processing efficiency in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology, 22*(5), 686-701.

Arnett, P. A. (2005). Longitudinal consistency of the relationship between depression symptoms and cognitive functioning in multiple sclerosis. *CNS spectrums, 10*(05), 372-382.

Arnett, P. A., Higginson, C. I., & Randolph, J. J. (2001). Depression in multiple sclerosis: relationship to planning ability. *Journal of the International Neuropsychological Society, 7*(6), 665-674.

Arnett, P. A., Higginson, C. I., Voss, W. D., Bender, W. I., Wurst, J. M., & Tippin, J. M. (1999). Depression in multiple sclerosis: relationship to working memory capacity. *Neuropsychology, 13*(4), 546.

Arnett, P. A., Higginson, C. I., Voss, W. D., Wright, B., Bender, W. I., Wurst, J. M., & Tippin, J. M. (1999). Depressed mood in multiple sclerosis: relationship to capacity-demanding memory and attentional functioning. *Neuropsychology, 13*(3), 434.

Arnett, P. A., Rao, S. M., Grafman, J., Bernardin, L., Luchetta, T., Binder, J. R., & Lobeck, L. (1997). Executive functions in multiple sclerosis: an analysis of temporal ordering, semantic encoding, and planning abilities. *Neuropsychology, 11*(4), 535.

Aupperle, R. L., Beatty, W. W., deNAP Shelton, F., & Gontkovsky, S. T. (2002). Three screening batteries to detect cognitive impairment in multiple sclerosis. *Multiple Sclerosis*, *8*(5), 382-389.

Bak, T., Rogers, T., Crawford, L., Hearn, V., Mathuranath, P., & Hodges, J. R. (2005). Cognitive bedside assessment in atypical parkinsonian syndromes. *Journal of Neurology, Neurosurgery & Psychiatry, 76*(3), 420-422.

Bak, T. H., & Mioshi, E. (2007). A cognitive bedside assessment beyond the MMSE: the Addenbrooke's Cognitive Examination. *Practical Neurology, 7*(4), 245-249.

Bak, T. H., O'Donovan, D. G., Xuereb, J. H., Boniface, S., & Hodges, J. R. (2001). Selective impairment of verb processing associated with pathological changes in Brodmann areas 44 and 45 in the motor neurone disease–dementia–aphasia syndrome. *Brain, 124*(1), 103-120.

Barbizet, J., & Cany, E. (1967). Clinical and psychometrical study of a patient with memory disturbances. *International Journal of Neurology, 7*(1), 44-54.

Barkhof, F. (2002). The clinico-radiological paradox in multiple sclerosis revisited. *Current Opinion in Neurology,* 15(3), 239-45.

Beatty, W. W., Goodkin, D. E., Hertsgaard, D., & Monson, N. (1990). Clinical and demographic predictors of cognitive performance in multiple sclerosis: Do diagnostic type, disease duration, and disability matter? *Archives of Neurology, 47*(3), 305-308.

Beatty, W. W., Goodkin, D. E., Monson, N., & Beatty, P. A. (1989). Cognitive disturbances in patients with relapsing remitting multiple sclerosis. *Archives of Neurology*, *46*(10), 1113-1119.

Beatty, W. W., Goodkin, D. E., Monson, N., Beatty, P. A., Hertsgaard, D. (1988). Anterograde and retrograde amnesia in patients with chronic progressive multiple sclerosis. *Archives of Neurology, 45,* 611-619.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*.

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review, 8*(1), 77-100.

Bell, B. D., Fine, J., Dow, C., Seidenberg, M., & Hermann, B. P. (2005). Temporal lobe epilepsy and the selective reminding test: the conventional 30-minute delay suffices. *Psychological assessment, 17*(1), 103.

Benedict, R. H., Morrow, S. A., Weinstock-Guttman, B., Cookfair, D., & Schretlen, D. J. (2010). Cognitive reserve moderates decline in information processing speed in multiple sclerosis patients. *Journal of the International Neuropsychological Society, 16*(05), 829-835.

Benedict, R. H., Weinstock-Guttman, B., Fishman, I., Sharma, J., Tjoa, C. W., & Bakshi, R. (2004). Prediction of neuropsychological impairment in multiple sclerosis: comparison of conventional magnetic resonance imaging measures of atrophy and lesion burden. *Archives of Neurology, 61*(2), 226-230.

Benedict, R. H., Shapiro, A., Priore, R., Miller, C., Munschauer, F., & Jacobs, L. (2000). Neuropsychological counseling improves social behavior in cognitively-impaired multiple sclerosis patients. *Multiple Sclerosis, 6*(6), 391-396.

Benedict, R. H. B., & Zivadinov, R. (2011). Risk factors for and management of cognitive dysfunction in multiple sclerosis. [10.1038/nrneurol.2011.61]. *Nat Rev Neurol, 7*(6), 332-342.

Bergendal, G., Fredrikson, S., & Almkvist, O. (2007). Selective decline in information processing in subgroups of multiple sclerosis: an 8-year longitudinal study. *European neurology, 57*(4), 193-202.

Bonello, P. J., Rapport, L. J., & Millis, S. R. (1997). Psychometric properties of the visual object and space perception battery in normal older adults. *The Clinical Neuropsychologist, 11*(4), 436-442.

Bonett, D. G., & Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika, 65*(1), 23-28.

Bonnet, M. C., Deloire, M. S., Salort, E., Dousset, V., Petry, K. G., Brochet, B., & Group, A. S. (2006). Evidence of cognitive compensation associated with educational level in early relapsing–remitting multiple sclerosis. *Journal of the neurological sciences, 251*(1), 23-28.

Borghi, M., Cavallo, M., Carletto, S., Ostacoli, L., Zuffranieri, M., Picci, R. L., . . . Bertolotto, A. (2013). Presence and significant determinants of cognitive impairment in a large sample of patients with multiple sclerosis. *PloS one, 8*(7), e69820.

Boringa, J., Lazeron, R. H., Reuling, I. E., Ader, H., Pfennings, L. E., Lindeboom, J., . . . Polman, C. (2001). The brief repeatable battery of neuropsychological tests: normative values allow application in multiple sclerosis clinical practice. *Multiple Sclerosis, 7*(4), 263-267.

Bschor, T., Kühl, K.-P., & Reischies, F. M. (2001). Spontaneous speech of patients with dementia of the Alzheimer type and mild cognitive impairment. *International Psychogeriatrics, 13*(03), 289-298.

Buschke, H., & Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology, 24*(11), 1019-1019.

Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods, 3*(1), 1-27.

Camp, S., Stevenson, V., Thompson, A., Ingle, G., Miller, D., Borras, C., . . . Filippi, M. (2005). A longitudinal study of cognition in primary progressive multiple sclerosis. *Brain, 128*(12), 2891-2898.

Camp, S., Stevenson, V., Thompson, A., Miller, D., Borras, C., Auriacombe, S., . . . Herisse-Dulo, L. (1999). Cognitive function in primary progressive and transitional progressive multiple sclerosis A controlled study with MRI correlates. *Brain, 122*(7), 1341-1348.

Canellopoulou, M., & Richardson, J. T. (1998). The role of executive function in imagery mnemonics: evidence from multiple sclerosis. *Neuropsychologia, 36*(11), 1181-1188.

Carone, D., Benedict, R., Fishman, I., & Weinstock-Guttman, B. (2005). Interpreting patient/informant discrepancies of reported cognitive symptoms in MS. *Journal of the International Neuropsychological Society, 11*(05), 574-583.

Carpenter, J. R., & Kenward, M. G. (2007). Missing data in randomised controlled trials-a practical guide. *London School of Hygiene*.

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research, 12*(1), 43-47.

Chelune, G., Stott, H., & Pinkston, J. (2008). Multiple sclerosis in textbook of clinical neuropsychology: Taylor and Francis, New York.

Chelune, G. J., Stott, H., & Pinkston, J. (2008). Multiple sclerosis. In J. E. M. J. H. Ricker (Ed.), *Textbook of clinical neuropsychology*. New York: Taylor & Francis.

Chen, I. (2012). More than meets the eye – the promises and pitfalls of MRI imaging in multiple sclerosis. *Multiple Sclerosis Discovery Forum,* News synthesis, 2012. (retrieved on 25/08/2016)

Chiaravalloti, N. D., Balzano, J., Moore, N. B., & DeLuca, J. (2009). The Open-Trial Selective Reminding Test (OT-SRT) as a tool for the assessment of learning and memory. *The Clinical Neuropsychologist, 23*(2), 231-254.

Chiaravalloti, N. D., & DeLuca, J. (2003). Assessing the behavioral consequences of multiple sclerosis: An application of the Frontal Systems Behavior Scale (FrSBe). *Cognitive and Behavioral Neurology, 16*(1), 54-67.

Chipchase, Y., Lincolns, N. B. (2001). Factors associated with carer strain in carers of people with multiple sclerosis. *Disability & Rehabilitation, 23*(17), 768-776.

Christodoulou, C., Melville, P., Scherl, W. F., Morgan, T., MacAllister, W. S., Canfora, D. M., . . . Krupp, L. B. (2005). Perceived cognitive dysfunction and observed neuropsychological performance: longitudinal relation in persons with multiple sclerosis. *Journal of the International Neuropsychological Society, 11*(05), 614-619.

Comi, G., Filippi, M., Martinelli, V., Campi, A., Rodegher, M., Alberoni, M., ... & Canal, N. (1995). Brain MRI correlates of cognitive impairment in primary and secondary progressive multiple sclerosis. *Journal of the Neurological Sciences*, *132*(2), 222-227.

Comparing hierarchical clustering dendrograms obtained by different distances & methods. (2013). Retrieved from http://stats.stackexchange.com/questions/63546/comparing-hierarchical-clustering-dendrograms-obtained-by-different-distances/63549#63549

Connick, P., Chandran, S., & Bak, T. H. (2013). Patterns of cognitive dysfunction in progressive MS. *Behavioural Neurology*, *27*(3), 259-265.

Connick, P., Kolappan, M., Bak, T. H., & Chandran, S. (2012). Verbal fluency as a rapid screening test for cognitive impairment in progressive multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 83*(3), 346-347.

Cosentino, S., & Stern, Y. (2005). Metacognitive theory and assessment in dementia: do we recognize our areas of weakness? *Journal of the International Neuropsychological Society, 11*(07), 910-919.

Croot, K., Hodges, J. R., & Patterson, K. (1999). Evidence for impaired sentence comprehension in early Alzheimer's disease. *Journal of the International Neuropsychological Society, 5*(05), 393-404.

Cutter, G. R., Baier, M. L., Rudick, R. A., Cookfair, D. L., Fischer, J. S., Petkau, J., . . . Confavreux, C. (1999). Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain, 122*(5), 871-882.

D'Amico, E., Leone, C., Hayrettin, T., & Patti, F. (2016). Can we define a rehabilitation strategy for cognitive impairment in progressive multiple sclerosis? A critical appraisal. *Multiple Sclerosis Journal*, *22*(5), 581-589.

D'Esposito, M., Onishi, K., Thompson, H., Robinson, K., Armstrong, C., & Grossman, M. (1996). Working memory impairments in multiple sclerosis: Evidence from a dual-task paradigm. *Neuropsychology, 10*(1), 51.

Dalgas, U., & Stenager, E. (2012). Exercise and disease progression in multiple sclerosis: can exercise slow down the progression of multiple sclerosis? *Therapeutic advances in neurological disorders, 5*(2), 81-95.

Danek, A. H., Gade, M., Lunardelli, A., & Rumiati, R. I. (2013). Tomato and Tuna: A Test for Language-Free Assessment of Action Understanding. *Cognitive and Behavioral Neurology*, *26*(4), 208-217.

de Groot, V., Beckerman, H., Uitdehaag, B. M., Hintzen, R. Q., Minneboo, A., Heymans, M. W., . . . Prognostication, F. (2009). Physical and cognitive functioning after 3 years can be predicted using information from the diagnostic process in recently diagnosed multiple sclerosis. *Archives of physical medicine and rehabilitation, 90*(9), 1478-1488.

De Sonneville, L., Boringa, J., Reuling, I., Lazeron, R., Ader, H., & Polman, C. (2002). Information processing characteristics in subtypes of multiple sclerosis. *Neuropsychologia, 40*(11), 1751-1765.

Deloire, M., Ruet, A., Hamel, D., Bonnet, M., & Brochet, B. (2010). Early cognitive impairment in multiple sclerosis predicts disability outcome several years later. *Multiple Sclerosis*.

DeLuca, J., Barbieri-Berger, S., & Johnson, S. K. (1994). The nature of memory impairments in multiple sclerosis: acquisition versus retrieval. *Journal of clinical and experimental Neuropsychology, 16*(2), 183-189.

DeLuca, J., Chelune, G. J., Tulsky, D. S., Lengenfelder, J., & Chiaravalloti, N. D. (2004). Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? *Journal of Clinical and Experimental Neuropsychology, 26*(4), 550-562.

DeLuca, J., Gaudino, E. A., Diamond, B. J., Christodoulou, C., & Engel, R. A. (1998). Acquisition and storage deficits in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *20*(3), 376-390.

Demaree, H. A., DeLuca, J., Gaudino, E. A., & Diamond, B. J. (1999). Speed of information processing as a key deficit in multiple sclerosis: implications for rehabilitation. *Journal of Neurology, Neurosurgery & Psychiatry, 67*(5), 661-663.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, 1-38.

Denney, D. R., Lynch, S. G., & Parmenter, B. A. (2008). A 3-year longitudinal study of cognitive impairment in patients with primary progressive multiple sclerosis: speed matters. *Journal of the neurological sciences, 267*(1), 129-136.

Denney, D. R., Sworowski, L. A., & Lynch, S. G. (2005). Cognitive impairment in three subtypes of multiple sclerosis. *Archives of Clinical Neuropsychology*, *20*(8), 967-981.

Denney, D. R., Lynch, S. G., Parmenter, B. A., & Horne, N. (2004). Cognitive impairment in relapsing and primary progressive multiple sclerosis: mostly a matter of speed. *Journal of the International Neuropsychological Society*, *10*(07), 948-956.

DeSousa, E. A., Albert, R. H., & Kalman, B. (2002). Cognitive impairments in multiple sclerosis: a review. *American Journal of Alzheimer's Disease and Other Dementias, 17*(1), 23-29.

Duque, B., Sepulcre, J., Bejarano, B., Samaranch, L., Pastor, P., & Villoslada, P. (2008). Memory decline evolves independently of disease activity in MS. *Multiple Sclerosis, 14*(7), 947-953.

Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression. I. *Biometrika, 37*(3-4), 409-428.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods, 4*(3), 272.

Filippi, M., Rocca, M., Benedict, R., DeLuca, J., Geurts, J., Rombouts, S., . . . Comi, G. (2010). The contribution of MRI in assessing cognitive impairment in multiple sclerosis. *Neurology, 75*(23), 2121-2128.

Filippi, M., Alberoni, M., Martinelli, V., Sirabian, G., Bressi, S., Canal, N., & Comi, G. (1994). Influence of clinical variables on neuropsychological performance in multiple sclerosis. *European Neurology*, *34*(6), 324-328.

Filley, C. M., Heaton, R. K., Nelson, L. M., Burks, J. S., & Franklin, G. M. (1989). A comparison of dementia in Alzheimer's disease and multiple sclerosis. *Archives of Neurology, 46*(2), 157-161.

Fischer, J. S. (2001). Cognitive impairment in multiple sclerosis. *Neurological Disease and Therapy, 53*, 233-256.

Fischer, J. S. (1988). Using the Wechsler Memory Scale-Revised to detect and characterize memory deficits in multiple sclerosis. *The Clinical Neuropsychologist*, *2*(2), 149-172.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*: MIT press.

Foong, J., Rozewicz, L., Chong, W., Thompson, A., Miller, D., & Ron, M. (2000). A comparison of neuropsychological deficits in primary and secondary progressive multiple sclerosis. *Journal of neurology, 247*(2), 97-101.

Foong, J., Rozewicz, L., Quaghebeur, G., Davie, C., Kartsounis, L., Thompson, A., . . . Ron, M. (1997). Executive functions in multiple sclerosis. *Brain, 120*, 15-26.

Foong, J., Rozewicz, L., Quaghebeur, G., Thompson, A., Miller, D., & Ron, M. (1998). Neuropsychological deficits in multiple sclerosis after acute relapse. *Journal of Neurology, Neurosurgery & Psychiatry, 64*(4), 529-532.

Frischer, J. M., Bramow, S., Dal-Bianco, A., Lucchinetti, C. F., Rauschka, H., Schmidbauer, M., . . . Lassmann, H. (2009). The relation between inflammation and neurodegeneration in multiple sclerosis brains. *Brain, 132*(5), 1175-1189.

Gaudino, E. A., Chiaravalloti, N. D., DeLuca, J., & Diamond, B. J. (2001). A comparison of memory performance in relapsing–remitting, primary progressive and secondary progressive, multiple sclerosis. *Cognitive and Behavioral Neurology, 14*(1), 32-44.

Giles, E., Patterson, K., & Hodges, J. R. (1996). Performance on the Boston Cookie Theft picture description task in patients with early dementia of the Alzheimer's type: missing information. *Aphasiology, 10*(4), 395-408.

Gladsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment, 6*(2), 147-178.

Glanz, B. I., Healy, B. C., Hviid, L. E., Chitnis, T., & Weiner, H. L. (2012). Cognitive deterioration in patients with early multiple sclerosis: a 5-year study. *Journal of Neurology, Neurosurgery & Psychiatry, 83*(1), 38-43.

Goodglass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*: Lea & Febiger.

Goodin, D. S. (2006). Magnetic resonance imaging as a surrogate outcome measure of disability in multiple sclerosis: have we been overly harsh in our assessment?. *Annals of Neurology*, *59*(4), 597-605.

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate behavioral research*, *26*(3), 499-510.

Grigsby, J., Kaye, K., & Busenbark, D. (1994). Alphanumeric sequencing: a report on a brief measure of information processing used among persons with multiple sclerosis. *Perceptual and motor skills, 78*(3), 883-887.

Gronwall, D. (1977). Paced auditory serial-addition task: a measure of recovery from concussion. *Perceptual and motor skills, 44*(2), 367-373.

Grossman, M., Armstrong, C., Onishi, K., Thompson, H., Schaefer, B., Robinson, K., . . . Rostami, A. (1994). Patterns of Cognitive Impairment in Relapsing-Remitting and Chronic Progressive Multiple Sclerosis. *Cognitive and Behavioral Neurology, 7*(3), 194-210.

Grossman, M., Robinson, K. M., Onishi, K., Thompson, H., Cohen, J., & D'Esposito, M. (1995). Sentence comprehension in multiple sclerosis. *Acta Neurologica Scandinavica, 92*(4), 324-331. doi: 10.1111/j.1600-0404.1995.tb00137.x

Haase, C., Tinnefeld, M., Daum, I., Ganz, R., Haupts, M., & Faustmann, P. (2004). Cognitive, but not mood dysfunction develops in multiple sclerosis during 7 years of follow-up. *European neurology, 52*(2), 92-95.

Hackmack, K., Weygandt, M., Wuerfel, J., Pfueller, C. F., Bellmann-Strobl, J., Paul, F., et al. (2012). Can we overcome the 'clinico-radiological paradox' in multiple sclerosis? *Journal of Neurology, 259*(10), 2151-60.

Halpern, D. F. (2013). *Sex differences in cognitive abilities*: Psychology press.

Harel, O., & Zhou, X. H. (2007). Multiple imputation: review of theory, implementation and software. *Statistics in Medicine*, *26*(16), 3057-3077.

Heaton, R. K., Nelson, L. M., Thompson, D. S., Burks, J. S., & Franklin, G. M. (1985). Neuropsychological findings in relapsing-remitting and chronic-progressive multiple sclerosis. *Journal of Consulting and Clinical Psychology*, *53*(1), 103-110.

Henry, J. D., & Beatty, W. W. (2006). Verbal fluency deficits in multiple sclerosis. *Neuropsychologia, 44*(7), 1166-1174.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology, 18*(2), 284.

Hierarchical or Two-step cluster analysis for binary data? (2014). Retrieved from http://stats.stackexchange.com/questions/116856/hierarchical-or-two-step-cluster-analysis-for-binary-data

Hobart, J., Lamping, D., Fitzpatrick, R., Riazi, A., & Thompson, A. (2001). The multiple sclerosis impact scale (MSIS-29) a new patient-based outcome measure. *Brain, 124*(5), 962-973.

Hoffmann, S., Tittgemeyer, M., & von Cramon, D. Y. (2007). Cognitive impairment in multiple sclerosis. *Current Opinion in Neurology, 20*(3), 275-280.

Honaker, J., King, G., & Blackwell, M. (2015). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1-47.

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, *54*(2), 561-581.

Huijbregts, S. C., Kalkers, N. F., de Sonneville, L. M., de Groot, V., & Polman, C. H. (2006). Cognitive impairment and decline in different MS subtypes. *Journal of the Neurological Sciences*, *245*(1), 187-194.

Hulst, E. J. (2012). *The Heterogeneity of Cognitive Impairment in Amyotrophic Lateral Sclerosis.* Doctor of Philosophy, The University of Edinburgh, Edinburgh.

Inglese M, Benedetti B, Filippi M. (2005). The relation between MRI measures of inflammation and neurodegeneration in multiple sclerosis. *Journal of the Neurological Sciences, 233(*1-2), 15-9.

Jennekens-Schinkel, A., Laboyrie, P., Lanser, J., & Van der Velde, E. (1990). Cognition in patients with multiple sclerosis: After four years. *Journal of the neurological sciences, 99*(2), 229-247.

Jongen, P., Ter Horst, A., & Brands, A. (2012). Cognitive impairment in multiple sclerosis. *Minerva medica, 103*(2), 73-96.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344): John Wiley & Sons.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, 347-361.

Kingwell, E., Marriott, J. J., Jetté, N., Pringsheim, T., Makhani, N., Morrow, S. A., . . . Kulaga, S. (2013). Incidence and prevalence of multiple sclerosis in Europe: a systematic review. *BMC neurology, 13*(1), 128.

Kujala, P., Portin, R., & Ruutiainen, J. (1997). The progress of cognitive decline in multiple sclerosis. *Brain, 120*, 289-297.

Kujala, P., Portin, R., Revonsuo, A., Ruutiainen, J. (1995). Attention related performance in two cognitively different subtypes of patients with multiple sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry, 59,* 77-82.

Kurtzke, J. F. (1970). Neurologic impairment in multiple sclerosis and the disability status scale. *Acta Neurologica Scandinavica, 46*(4-5), 493-512.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis an expanded disability status scale (EDSS). *Neurology, 33*(11), 1444-1444.

Langdon, D. W., Amato, M. P., Boringa, J., Brochet, B., Foley, F., Fredrikson, S., ... & Reder, A. T. (2012). Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). *Multiple Sclerosis Journal*, *18*(6), 891-898.

Langdon, D. W. (2011). Cognition in multiple sclerosis. *Current opinion in neurology, 24*(3), 244-249.

Larner, A. (2007). Addenbrooke's Cognitive Examination-Revised (ACE-R) in day-to-day clinical practice. *Age and ageing, 36*(6), 685-686.

Larrabee, G. J., Trahan, D. E., & Levin, H. S. (2000). Normative data for a six-trial administration of the Verbal Selective Reminding Test. *The Clinical Neuropsychologist, 14*(1), 110-118.

Lazeron, R. H., Boringa, J., Schouten, M., Uitdehaag, B. M., Bergers, E., Lindeboom, J., . . . Polman, C. (2005). Brain atrophy and lesion load as explaining parameters for cognitive impairment in multiple sclerosis. *Multiple Sclerosis, 11*(5), 524-531.

Lengenfelder, J., Bryant, D., Diamond, B. J., Kalmar, J. H., Moore, N. B., & DeLuca, J. (2006). Processing speed interacts with working memory efficiency in multiple sclerosis. *Archives of Clinical Neuropsychology, 21*(3), 229-238.

Lezak, M. D. (2004). *Neuropsychological assessment*: Oxford university press.

Little, R. J., & Rubin, D. B. (2002). Single imputation methods. *Statistical Analysis with Missing Data, Second Edition*, 59-74.

Litvan, I., Grafman, J., Vendrell, P., & Martinez, J. M. (1988). Slowed information processing in multiple sclerosis. *Archives of Neurology, 45*(3), 281-285.

Litvan, I., Grafman, J., Vendrell, P., Martinez, J. M., Junqué, C., Vendrell, J. M., & Barraquer-Bordas, J. L. (1988). Multiple memory deficits in patients with multiple sclerosis: Exploring the working memory system. *Archives of Neurology*, *45*(6), 607-610.

Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sørensen, P. S., Thompson, A. J., . . . Barkhof, F. (2014). Defining the clinical course of multiple sclerosis The 2013 revisions. *Neurology, 83*(3), 278-286.

Lynch, S. G., Parmenter, B. A., & Denney, D. R. (2005). The association between cognitive impairment and physical disability in multiple sclerosis. *Multiple Sclerosis, 11*(4), 469-476.

MacAllister, W., Belman, A., Milazzo, M., Weisbrot, D., Christodoulou, C., Scherl, W., . . . Krupp, L. (2005). Cognitive functioning in children and adolescents with multiple sclerosis. *Neurology, 64*(8), 1422-1425.

Mackenzie, I., Morant, S., Bloomfield, G., MacDonald, T., & O'riordan, J. (2014). Incidence and prevalence of multiple sclerosis in the UK 1990–2010: a descriptive study in the General Practice Research Database. *Journal of Neurology, Neurosurgery & Psychiatry, 85*(1), 76-84.

Mackenzie, C., & Green, J. (2009). Cognitive–linguistic deficit and speech intelligibility in chronic progressive multiple sclerosis. *International Journal of Language & Communication Disorders*, *44*(4), 401-420.

Mariani, C., Farina, E., Cappa, S., Anzola, G., Faglia, L., Bevilacqua, L., . . . Vignolo, L. (1991). Neuropsychological assessment in multiple sclerosis: a follow-up study with magnetic resonance imaging. *Journal of neurology, 238*(7), 395-400.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology, 55*(11), 1613–1620.

McCarthy, R. A., & Warrington, E. K. (1990). *Cognitive neuropsychology: A clinical introduction*: Academic press.

McKenna, P., & Warrington, E. K. (1983). *Graded naming test: Manual*: NFER-Nelson.

Meier, D. S., Weiner, H. L., & Guttmann, C. R. (2007). MR imaging intensity modeling of damage and repair in multiple sclerosis: relationship of short-term lesion recovery to progression and disability. *American Journal of Neuroradiology*, *28*(10), 1956-1963.

Miller, A. E. (2001). Clinical features. In S. D. Cook (Ed.), *Handbook of multiple sclerosis* (3 ed.). New York: Marcel Dekker.

Miller, E. M. (1994). Intelligence and brain myelination: A hypothesis. *Personality and Individual Differences*, *17*(6), 803-832.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45*(3), 325-342.

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*(2), 159-179.

Milo, R., & Kahana, E. (2010). Multiple sclerosis: geoepidemiology, genetics and the environment. *Autoimmunity reviews, 9*(5), A387-A394.

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry, 21*(11), 1078-1085.

Mollison, D., Sellar, R., Bastin, M., Mollison, D., Chandran, S., Wardlaw, J., Connick, P. (under review). The clinico-radiological paradox of cognitive function and MRI burden of white matter lesions in people with multiple sclerosis: a systematic review and meta-analysis. *Multiple Sclerosis Journal*

Motl, R. W., & Pilutti, L. A. (2012). The benefits of exercise training in multiple sclerosis. *Nature Reviews Neurology, 8*(9), 487-497.

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART)*: Nfer-Nelson.

Noseworthy, J., Lucchinetti, C., Rodriguez, M., & Weinshenker, B. (2000). Medical progress. Multiple Sclerosis. *The New England Journal of Medicine, 343*(13), 938-952.

O'Keeffe, F. M., Murray, B., Coen, R. F., Dockree, P. M., Bellgrove, M. A., Garavan, H., ... & Robertson, I. H. (2007). Loss of insight in frontotemporal dementia, corticobasal degeneration and progressive supranuclear palsy. *Brain*, *130*(3), 753-764.

OECD. (2014). *Annual Scottish Labour Force Survey*

Parmenter, B., Weinstock-Guttman, B., Garg, N., Munschauer, F., & Benedict, R. H. (2007). Screening for cognitive impairment in multiple sclerosis using the Symbol Digit Modalities Test. *Multiple Sclerosis, 13*(1), 52-57.

Patti, F., Failla, G., Ciancio, M., L'Episcopo, M., & Reggio, A. (1998). Neuropsychological, neuroradiological and clinical findings in multiple sclerosis. A 3 year follow-up study. *European Journal of Neurology, 5*(3), 283-286.

Paul, R. H., Beatty, W. W., Schneider, R., Blanco, C., & Hames, K. (1998). Impairments of attention in individuals with multiple sclerosis. *Multiple Sclerosis, 4*(5), 433-439.

Penke, L., Maniega, S. M., Bastin, M., Hernandez, M.V., Murray, C., Royle, N., et al. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry, 17*(10), 1026–30.

Penke, L., Maniega, S. M., Murray, C., Gow, A. J., Hernández, M. V., Clayden, J. D., et al. (2010). A general factor of brain white matter integrity predicts information processing speed in healthy older people. *The Journal of Neuroscience, 30*(22), 7569–74.

Peters, A. (1996). Age-related changes in oligodendrocytes in monkey cerebral cortex. *Journal of Comparative Neurology*, *371*(1), 153-163.

Pettit, L. D. (2015). White matter integrity, executive dysfunction, and processing speed in amyotrophic lateral sclerosis [Unpublished doctoral dissertation]. [Edinburgh]: University of Edinburgh

Peyser, J. M., Rao, S. M., LaRocca, N. G., & Kaplan, E. (1990). Guidelines for neuropsychological research in multiple sclerosis. *Archives of Neurology, 47*(1), 94-97.

Piccinin, A. M., Rabbitt, P. (1999). Contribution of cognitive abilities to performance and improvement on a substitution coding task. *Psychology and Aging, 14*(4), 539.

Piras, M. R., Magnano, I., Canu, E. D. G., Paulus, K. S., Satta, W. M., Soddu, A., . . . Aiello, I. (2003). Longitudinal study of cognitive dysfunction in multiple sclerosis: neuropsychological,

neuroradiological, and neurophysiological findings. *Journal of Neurology, Neurosurgery & Psychiatry, 74*(7), 878-885.

Planche, V., Gibelin, M., Cregut, D., Pereira, B., Clavelou, P. (2015). Cognitive impairment in a population-based study of patients with multiple sclerosis: Differences between late relapsing-remitting, secondary progressive and primary progressive multiple sclerosis. *European Journal of Neurology, 23, 282–289.*

Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., . . . Kappos, L. (2011). Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of Neurology, 69*(2), 292-302.

Potagas, C., Giogkaraki, E., Koutsis, G., Mandellos, D., Tsirempolou, E., Sfagos, C., & Vassilopoulos, D. (2008). Cognitive impairment in different MS subtypes and clinically isolated syndromes. *Journal of the Neurological Sciences*, *267*(1), 100-106.

Poser, S., Kurtzke, J. F., Poser, W., & Schlaf, G. (1989). Survival in multiple sclerosis. *Journal of Clinical Epidemiology, 42*(2), 159-168.

Prakash, R., Snook, E., Lewis, J., Motl, R., & Kramer, A. (2008). Cognitive impairments in relapsing-remitting multiple sclerosis: a meta-analysis. *Multiple Sclerosis, 14*(9), 1250-1261.

Prakash, R. S., Snook, E. M., Motl, R. W., & Kramer, A. F. (2010). Aerobic fitness is associated with gray matter volume and white matter integrity in multiple sclerosis. *Brain Research, 1341*, 41-51.

Pryse-Phillips, W., Costello, F. (2001). Epidemiology of multiple sclerosis. In S. D. Cook (Ed.), *Handbook of Multiple Sclerosis* (3 ed.). New York: Marcel Dekker.

Rao, S. (1990). A manual for the brief, repeatable battery of neuropsychological tests in multiple sclerosis. *New York: National Multiple Sclerosis Society*, 121-123.

Rao, S. M., Aubin-Faubert, P. S., & Leo, G. J. (1989). Information processing speed in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology, 11*(4), 471-477.

Rao, S. M., Leo, G. J., Bernardin, L., & Unverzagt, F. (1991). Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology, 41*(5), 685-691.

Reichenberg, A., Yirmiya, R., Schuld, A., Kraus, T., Haack, M., Morag, A., & Pollmächer, T. (2001). Cytokine-associated emotional and cognitive disturbances in humans. *Archives of General Psychiatry*, *58*(5), 445-452.

Reuter, F., Zaaraoui, W., Crespy, L., Faivre, A., Rico, A., Malikova, I., . . . Pelletier, J. (2010). Frequency of cognitive impairment dramatically increases during the first 5 years of multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry*, jnnp. 2010.213744.

Rewaj, P. J. (2013). *The Nature of Language Impairment in Motor Neurone Disease.* Doctor of Philosophy, The University of Edinburgh, Edinburgh.

Reynolds, R., Roncaroli, F., Nicholas, R., Radotra, B., Gveric, D., & Howell, O. (2011). The neuropathological basis of clinical progression in multiple sclerosis. *Acta Neuropathologica*, *122*(2), 155-170.

Riazi, A., Hobart, J., Lamping, D., Fitzpatrick, R., & Thompson, A. (2002). Multiple Sclerosis Impact Scale (MSIS-29): reliability and validity in hospital based samples. *Journal of Neurology, Neurosurgery & Psychiatry, 73*(6), 701-704.

Riddoch, M. J., & Humphreys, G. W. (1993). *BORB: Birmingham object recognition battery*: LEA.

Robinson, L., Gemski, A., Abley, C., Bond, J., Keady, J., Campbell, S., . . . Manthorpe, J. (2011). The transition to dementia–individual and family experiences of receiving a diagnosis: a review. *International Psychogeriatrics, 23*(07), 1026-1043.

Ron, M., Callanan, M., & Warrington, E. (1991). Cognitive abnormalities in multiple sclerosis: a psychometric and MRI study. *Psychological medicine, 21*(01), 59-68.

Ross, T. P. (2003). The reliability of cluster and switch scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology, 18*(2), 153-164.

Ruet, A., Deloire, M., Charré-Morin, J., Hamel, D., Brochet, B. (2013). Cognitive impairment differs between primary progressive and relapsing-remitting MS. *Neurology, 80(16),* 1501–1508.

Ryan, L., Clark, C. M., Klonoff, H., Li, D., & Paty, D. (1996). Patterns of cognitive impairment in relapsing–remitting multiple sclerosis and their relationship to neuropathology on magnetic resonance images. *Neuropsychology, 10*(2), 176.

Sadovnick, A., Ebers, G., Wilson, R., & Paty, D. (1992). Life expectancy in patients attending multiple sclerosis clinics. *Neurology, 42*(5), 991-991.

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*, *33*(4), 545-571.

Schoo, L. A., van Zandvoort, M. J., Biessels, G., Kappelle, L. J., & Postma, A. (2013). Insight in Cognition: Self-Awareness of performance across cognitive domains. *Applied Neuropsychology: Adult, 20*(2), 95-102.

Schwid, S. R., Goodman, A. D., Weinstein, A., McDermott, M. P., Johnson, K. P., & Group, C. S. (2007). Cognitive function in relapsing multiple sclerosis: minimal changes in a 10-year clinical trial. *Journal of the Neurological Sciences, 255*(1), 57-63.

Scotland, J. L., Whittle, I. R., Deary, I. J. (2012). Cognitive functioning in newly presenting patients with supratentorial intracranial tumors: is there a role for inspection time? *Neuro-oncology*, nor222.

Seewann, A., Kooi, E. J., Roosendaal, S. D., Barkhof, F., van der Valk, P., Geurts, J. J. (2009). Translating pathology in multiple sclerosis: the combination of postmortem imaging, histopathology and clinical findings. *Acta Neurologica Scandinavica, 119*(6), 349-55.

Sheridan, L. K., Fitzgerald, H. E., Adams, K. M., Nigg, J. T., Martel, M. M., Puttler, L. I., . . . Zucker, R. A. (2006). Normative Symbol Digit Modalities Test performance in a community-based sample. *Archives of Clinical Neuropsychology, 21*(1), 23-28.

Shipley, B. A., Deary, I. J., Tan, J., Christie, G., Starr, J. M. (2002). Efficiency of temporal order discrimination as an indicator of bradyphrenia in Parkinson's disease: the inspection time loop task. *Neuropsychologia, 40*(8), 1488–93.

Similarity Coefficients for binary data: Why choose Jaccard over Russell and Rao? (2013). Retrieved from http://www.statistics4u.com/fundstat_eng/cc_distance_meas.html

Simioni, S., Ruffieux, C., Kleeberg, J., Bruggimann, L., Du Pasquier, R. A., ANNONI, J., & Schluep, M. (2009). Progressive decline of decision-making performances during multiple sclerosis. *Journal of the International Neuropsychological Society, 15*(02), 291-295.

Smith, A. (2002). *Symbol digit modalities test: Manual*: Western Psychological Corporation.

Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy. The principles and practice of numerical classification*.

Sperling, R. A., Guttmann, C. R., Hohol, M. J., Warfield, S. K., Jakab, M., Parente, M., . . . Orav, E. J. (2001). Regional magnetic resonance imaging lesion burden and cognitive function in multiple sclerosis: a longitudinal study. *Archives of Neurology, 58*(1), 115-121.

Stadelmann, C., Wegner, C., & Brück, W. (2011). Inflammation, demyelination, and degeneration—recent insights from MS pathology. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1812*(2), 275-282.

Stadelmann, C. (2011). Multiple sclerosis as a neurodegenerative disease: pathology, mechanisms and therapeutic implications. *Current Opinion in Neurology*, *24*(3), 224-229.

Staff, N. P., Lucchinetti, C. F., & Keegan, B. M. (2009). Multiple sclerosis with predominant, severe cognitive impairment. *Archives of Neurology, 66*(9), 1139-1143.

Stebbins, G. T., Wilson, R. S., Gilley, D. W., Bernard, B. A., & Fox, J. H. (1990). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *The Clinical Neuropsychologist, 4*(1), 18-24.

Strober, L. B., Rao, S. M., Lee, J.-C., Fischer, E., & Rudick, R. (2014). Cognitive impairment in multiple sclerosis: An 18 year follow-up study. *Multiple sclerosis and related disorders, 3*(4), 473-481.

Stroud, N. M., & Minahan, C. L. (2009). The impact of regular physical activity on fatigue, depression and quality of life in persons with multiple sclerosis. *Health Qual Life Outcomes, 7*(1), 68.

Sumowski, J., Wylie, G., Gonnella, A., Chiaravalloti, N., & Deluca, J. (2010). Premorbid cognitive leisure independently contributes to cognitive reserve in multiple sclerosis. *Neurology, 75*(16), 1428-1431.

Sumowski, J. F., Chiaravalloti, N., Wylie, G., & DeLuca, J. (2009). Cognitive reserve moderates the negative effect of brain atrophy on cognitive efficiency in multiple sclerosis. *Journal of the International Neuropsychological Society, 15*(04), 606-612.

Sumowski, J. F., Rocca, M. A., Leavitt, V. M., Riccitelli, G., Comi, G., DeLuca, J., & Filippi, M. (2013). Brain reserve and cognitive reserve in multiple sclerosis What you've got and how you use it. *Neurology, 80*(24), 2186-2193.

Thompson, A., Montalban, X., Barkhof, F., Brochet, B., Filippi, M., Miller, D., . . . McDonald, W. (2000). Diagnostic criteria for primary progressive multiple sclerosis: a position paper. *Annals of neurology, 47*(6), 831-835.

Thompson, A. J., Kermode, A. G., MacManus, D. G., Kendall, B. E., Kingsley, D. P., Moseley, I. F., & McDonald, W. I. (1990). Patterns of disease activity in multiple sclerosis: clinical and magnetic resonance imaging study. *Bmj*, *300*(6725), 631-634.

Thornton, A. E., & Raz, N. (1997). Memory impairment in multiple sclerosis: a quantitative review. *Neuropsychology, 11*(3), 357.

Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology, 14*(2), 167-177.

Trapp, B. D., & Nave, K.-A. (2008). Multiple sclerosis: an immune or neurodegenerative disorder? *Annu. Rev. Neurosci., 31*, 247-269.

Turner, A. P., Kivlahan, D. R., & Haselkorn, J. K. (2009). Exercise and Quality of Life Among People With Multiple Sclerosis: Looking Beyond Physical Functioning to Mental Health and Participation in Life. *Archives of Physical Medicine and Rehabilitation, 90*(3), 420-428. doi: http://dx.doi.org/10.1016/j.apmr.2008.09.558

Velikonja, O., Čurić, K., Ožura, A., & Jazbec, S. Š. (2010). Influence of sports climbing and yoga on spasticity, cognitive function, mood and fatigue in patients with multiple sclerosis. *Clinical Neurology and Neurosurgery, 112*(7), 597-601. doi: http://dx.doi.org/10.1016/j.clineuro.2010.03.006

Videbech, P., & Ravnkilde, B. (2004). Hippocampal volume and depression: a meta-analysis of MRI studies. *American Journal of Psychiatry, 161*(11), 1957-1966.

Vleugels, L., Lafosse, C., n van Nunen, A., Nachtergaele, S., Ketelaer, P., Charlier, M., & Vandenbussche, E. (2000). Visuoperceptual impairment in multiple sclerosis patients diagnosed with neuropsychological tasks. *Multiple Sclerosis, 6*(4), 241-254.

Wallace, G. L., & Holmes, S. (1993). Cognitive-linguistic assessment of individuals with multiple sclerosis. *Multiple Sclerosis, 1*(40), 4.

Warrington, E. K. (1997). The graded naming test: A restandardisation. *Neuropsychological Rehabilitation, 7*(2), 143-146.

Warrington, E. K., & James, M. (1991). *The visual object and space perception battery*: Thames Valley Test Company Bury St Edmunds.

WMA. (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama, 310*(20), 2191.

Yokoi, T., & Okamura, H. (2013). Why do dementia patients become unable to lead a daily life with decreasing cognitive function? *Dementia, 12*(5), 551-568.

Zivadinov, R., Sepcic, J., Nasuelli, D., De Masi, R., Bragadin, L. M., Tommasi, M., . . . Ukmar, M. (2001). A longitudinal study of brain atrophy and cognitive disturbances in the early phase of relapsing-remitting multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry, 70*(6), 773-780.