The University of Leicester

Screening for Cognitive Impairments in People with Multiple Sclerosis

Thesis submitted as part fulfilment of the requirements for the degree of Doctorate in Clinical Psychology

by

Richard M. Burton

M.Sc., B.Sc.(Hons)

May 2012

Table of Contents

Declaration

This thesis is principally the work of the author, Richard M. Burton, and no part has been submitted for any other academic award at this or any other educational establishment.

Signed .....

Richard M. Burton

Acknowledgements

I would like to thank the participants who were involved in this study and the staff of both the Neurology Outpatients and Neuropsychology departments where the research was carried out for their invaluable involvement in the project. I would also like to thank my supervisors Prof. Mike Wang and Dr. Joanne Kirkland for their advice and expertise while I have been completing the research. In addition I would like to thank Eric Gardner, Margaret Pepe and her research student Aastha Bansal for their invaluable advice regarding statistical tests.

Word Count

	Main Text	Tables	References
Literature Review	8,190	797	1,072
Research Report	12,579	933	2,466
Critical Appraisal	4,246		
Appendices (Excluding mandatory appendices)	4,060		
Total Word Count for Thesis	29,075		

Abstract

Richard Burton

Screening for Cognitive Problems in People with Multiple Sclerosis

Multiple Sclerosis is the most commonly found neurological condition among young adults and in early adulthood. Symptoms exhibit 'disease heterogeneity' and vary from sufferer to sufferer. Cognitive impairments in the domains of processing speed, memory, attention, visuospatial perception and executive function are commonly found.

Literature Review: A systematic review of the literature on screening for cognitive impairments is presented. The paper considers the key measures of validity and reliability for the screening tests reviewed and uses a standard appraisal tool to assess the methodology of each study. The review concludes that, of the wide variety of tools, which will be best to use is dependent upon the needs of the client and the resources of the service assessing them.

Research Report: The study examined the combined validity of two short screening tests – the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) and Symbol Digit Modalities Test (SDMT). Logistic regression was used to generate a combined score for the two tests which was compared to the Minimal Assessment of Cognitive Function In Multiple Sclerosis (MACFIMS) psychometric battery as a 'gold standard' measure and Receiver Operating Characteristic (ROC) analysis carried out. While the sample was underpowered, the logistic regression method produced superior area under the curve for the combined scores. Clinical implications of the results, recommendations based on the findings and potential future research projects were discussed.

Critical Appraisal: A reflective account of the process of carrying out the research and what was learned from it is provided.

List of Tables

Literature Review

Research Report

List of Figures

Research Report

Literature Review

Screening for Cognitive Dysfunction in Multiple Sclerosis: A Critical Review of the Literature

Target Journal: Journal of Neurology

(See Appendix W)

Screening for Cognitive Dysfunction in Multiple Sclerosis: A Critical Review of the Literature

Abstract

Cognitive impairment is a common symptom of the demyelinating central nervous system condition Multiple Sclerosis (Rao et al., 1991a). Such impairments can be assessed clinically using a battery of neuropsychological tests which are time intensive. By using screening tests, patients needing such assessment can be more efficiently identified. Scherer (2007) has suggested features valuable for evaluating screening tests. The current review sought to ascertain the present state of research into valid screening tests and compare them using Scherer's (2007) criteria. The existing literature was searched using PsychINFO, Scopus, PubMed and Web of Knowledge identifying 112 papers which were filtered down to 13 for evaluation covering 11 screening tests. The relative merits of these were tabulated for comparison. The best quoted levels of accuracy in the literature were for the Faces Symbol Test (FST; Scherer et al., 2007) and the Symbol Digit Modalities Test (SDMT; Parmenter et al., 2007). However, papers authored by O'Brien et al. (2007) and Lechner-Scott et al. (2010) obtained the best results in the assessment of study quality Which screening tool is used should depend on the situation of the client and the service assessing them.

Introduction

Multiple Sclerosis (MS) is the most frequently found neurological condition among young adults and in middle adulthood. The majority of symptoms first appear between the ages of 20 and 40 years of age. MS is an autoimmune disease since the neurological damage typical of the illness appears to be caused by the action of the body's immune system. Physical symptoms include fatigue, visual problems, muscle spasms and spasticity, musculoskeletal or neuropathic pain, bowel or bladder problems and mobility difficulties. These are measured clinically using the Expanded Disability Status Scale (EDSS; Kurtzke, 1983) which gives a rating of severity between 0 and 10 (where 0 is no disability and 10 is death from MS). MS also produces a variety of cognitive and emotional difficulties that include: depression, anxiety, problems with attention or memory and reduced mental processing speed.

Rosati (2001), in a review of the global prevalence rates of Multiple Sclerosis, cited figures in the UK of between 74 and 112 cases per 100,000 population in England and Wales. Rates are higher in Scotland at between 145 and 193 per 100,000. An increasing prevalence among women has been found by Hirst et al. (2009) and Sellner et al. (2011). National Institute for Health and Clinical Excellence (2003) estimate that, in England and Wales, between 1800 and 3400 people are newly diagnosed with MS every year and about 52,000 to 62,000 people have MS.

Guidelines for the management of MS in the NHS published by the National Institute for Health and Clinical Excellence (NICE, 2003) state that a specialist neurological team and neurological rehabilitation service should be available to everyone who has MS when they need this. These teams should have, as a minimum, doctors, nurses, physiotherapists, occupational therapists, speech and language therapists, clinical psychologists and social workers. Given the range of physical and

cognitive difficulties found in MS, sufferers may need support from a range of these professionals at various times so it is a key priority that information about people with MS is shared effectively between services so that provision is seamless. Services are expected to be responsive to the needs of patients and there should be an easy way of contacting the team – for example through an MS specialist nurse.

While it is commonly considered a discrete condition, there are four main types of MS diagnosis. The most common is the Relapsing-Remitting type affecting around half of all sufferers and characterised by periods where symptoms are evident, interspersed with periods of remission where functioning becomes more normal. No progression of the illness is evident in this diagnosis but if a more permanent progression is seen at a later stage, this is then diagnosed as the Secondary-Progressive type (the next most common diagnosis). A diagnosis of Primary-Progressive MS is given when the illness is most aggressive and there is no remission period, only a gradually increasing level of deterioration from disease onset. The least common type is Progressive Relapsing MS, which has a progressive nature but is punctuated by periods of acute relapse which may or may not remit.

Treatment approaches for MS fall into three categories – treating MS relapses; treating specific symptoms; and treatments to slow the progression of the illness. There is currently no cure for the disease. When relapses occur they are treated using steroids such as methylprednisolone to hasten recovery. Medical specialists in other parts of the NHS are used to treat specific symptoms. For example physiotherapists may aid with musculoskeletal pain or spasticity and clinical psychologists or neuropsychologists provide assessment and treatment of cognitive difficulties. To slow progression of the illness medications which suppress the immune system are used. These are typically

administered by injection under the skin or into a vein. Typical medications include natulizumab, interferon-beta and glatiramer acetate.

A seminal investigation into the cognitive deficits associated with MS was carried out by Rao, Leo, Bernardin and Unverzagt (1991a). Comparing 100 MS sufferers with 100 demographically matched and healthy controls on a range of 31 test indices, they found the MS group were more frequently impaired on measures of recent memory, sustained attention, verbal fluency, conceptual reasoning and visuospatial perception. Frequency of cognitive impairments in MS sufferers was, however, lower than had previously been thought at around 43%. Other studies have found rates of incidence which vary between 40-70%, although there is an increasing consensus quoting a level at approximately 50%.

Testing of the cognitive deficits associated with MS is usually carried out by a qualified neuropsychologist and therefore requires a referral from an MS clinic to a separate neuropsychology team. A range of psychometric tests are available to neuropsychologists to assess a sufferer's performance in the cognitive domains outlined above. Tests usually assess specific domains – for example the Paced Auditory Serial Addition Test (PASAT; Gronwall, 1977) examines working memory, attention and processing speed, while the California Verbal Learning Test - II (CVLT-II; Delis, Kramer, Kaplan & Ober, 2000) examines verbal memory performance.

Given the wide range of impairments, identification of cognitive problems from a neuropsychological perspective often requires a very wide, time consuming and expensive battery of psychometric tests. As a result, shorter screening tests are desirable in order to identify sufferers most in need of this in depth testing. Engel, Greim and Zettl (2007) outline the major neuropsychological tests used to identify cognitive deficits in MS sufferers. They suggest that they can be considered as detecting deficits

in three major areas of function: memory, attention and executive function, and suggest that effective testing of MS patients for cognitive problems should take into account the impact of both fatigue and depression on their cognitive abilities. The Beck Depression Inventory and the Fatigue Impact Scale are proposed as the best tests to be used for this purpose.

Scherer (2007) gave an overview of the screening tools evaluated in the published research and outlined the properties a good tool should possess. Each test has statistical properties that describe its ability to categorise patients as either impaired or cognitively intact. Test sensitivity is defined as the ability of the test to correctly designate an impaired patient as test positive and so is associated with few false negative results. The opposite applies for the definition of test specificity: this is defined as the ability of the test to correctly assign a patient to a cognitively intact category and is associated with few false positive results. Other statistical properties are derived from these, such as positive and negative predictive values – probabilities that those with either positive or negative results on the test are respectively impaired or intact. Scherer (2007) also discussed test accuracy, defined as the proportion of all tested persons where the test gave the correct result and suggests the best measure of test accuracy is made by plotting a Receiver Operating Characteristics (ROC) curve. Every test has to have a designated cut-off point or threshold that designates either a positive or negative result. A ROC curve is obtained by plotting how the true positive rate and false positive rate vary as this threshold is changed. The area under the ROC curve (AUROC) is a measure of the test's accuracy. The closer the AUROC curve is to 1.0, the better the test is performing, while an area of 0.5 would designate a performance no better than chance. To derive these values, a screening test must be compared to an established

‘gold standard’ measure of cognitive impairment such as a more extensive battery of tests.

The current review was a systematic literature review to examine the state of current research into useful screening tests for cognitive impairment in MS. The characteristics of the tests which Scherer (2007) lists as important are outlined and areas where further research would be possible are suggested. All studies considered used a battery of neuropsychological tests as their ‘gold standard’ of comparison against which the screening test was compared.

Methodology

In September 2011 searches were made of six research databases. An initial search using the terms: “Multiple Sclerosis” and “Cognitive Screening” was carried out on PsychINFO, PsychEXTRA and PsychARTICLES, collectively. These databases were selected for their coverage of areas relevant to Neuropsychology. An additional search on this set of databases was also carried out using the alternative terms: “Multiple Sclerosis”, “Cognitive Dysfunction” and “Screening”. These latter search terms were generated from the keywords associated with articles produced by the initial search. In both searches, the databases were restricted to provide results of papers published in peer reviewed journals and where the search terms occurred in any field. Both searches returned a total of 10 results. After brief consideration of the results returned, it was decided that the latter set of terms would only be applied to the other three databases to search, since this had returned a better quality of results. Since the topic under consideration was also related to medicine, the PubMed database was searched. Searching for the terms above in either the abstract or title field yielded a total of 20 results. Finally, the same searches were applied to databases with wide breadth of coverage: Scopus and Web of Knowledge were selected for this purpose. A total of 30 results were returned from Scopus and 42 from Web of Knowledge.

The 102 results obtained across the database searches were then combined and any duplication of papers found in the results was eliminated, leaving only 70 articles. This set of papers was then screened by the researcher by reading abstracts to eliminate any that did not meet a specific set of criteria: that the papers should be available in English; must specifically focus on Multiple Sclerosis and not any other similar or associated syndromes; that papers must concern cognitive testing; and must be a

published paper and not a letter or correction. Application of these criteria reduced the sample to 36 articles for review (see Appendix A).

A further review of the articles using thorough reading of abstracts and the reports themselves was carried out with stricter criteria applied. These new criteria required papers to be: specifically about cognitive impairment in MS; original research and not reviews; and for papers to explicitly use neuropsychological tests or cognitive screening tests in their methodology rather than other methods such as MRI scanning (See Appendix B). These stricter criteria reduced the number of papers selected for review to 13. Details of why each of the final papers were either included or excluded from the review are provided in Appendix C.

Once the final set of 13 papers had been selected, essential information about each was obtained using a data extraction form and all papers were quality assessed using a standardised tool developed by the Public Health Resource Unit (2006). This tool was based upon a paper by Jaesche, Guyatt and Sackett (1994) on how to assess diagnostic tests and is freely available in the public domain. Copies of the assessment tool and the data extraction form can be found in Appendices D and E. A quality score was derived from the answers given to the questions on the assessment tool such that their quality could be compared. For questions on the tool requiring answers of 'yes', 'no' or 'can't tell' a score of one was given for a 'yes' answer, half for 'can't tell' and zero for 'no' except for question 4 where this scoring was reversed because a 'no' answer on that question was an indicator of quality. In the case of questions 7 and 8 scores could not be generated using this method. For question seven, half a point was given if there were quoted values for sensitivity and specificity, and a further half point for quoted accuracy figures. For question eight, half a point was scored if there were quoted levels of statistical significance given for the study results, and a further half a

point was given if confidence limits were also provided on quoted results. For question twelve no scoring was given as this question required qualitative, non-comparable answers to be given. Each study, therefore, obtained a quality score of between zero and eleven.

Results

Details of the quality appraisal scores, samples, methodology and findings of each of the studies considered are provided in the table below:

Table 1: Summary of results and findings from studies selected for the review of the literature

Study	Quality Appraisal Score	Screening Test	Gold Standard	Sample Size	Types of MS Diagnosis	Gender	Ethnicity	Mean Age(SD) in Years	Mean years of education	Mean duration since diagnosis	Sensitivity	Specificity	Accuracy
Beatty, Paul, Wilbanks, Hames, Blanco and Goodkin (1995)	7.0	Screening Examination for Cognitive Impairment (SEFCI)	Researcher selected 2-hour battery	103 MS 32 HC	Not Given	68.9% Female, 31.1% Male	Not Given	45.2 (9.3)	14.5 (2.4)	10.0 (6.8)	0.86	0.90	Not given
Basso, Beason-Hazen, Lynn, Rammohan and Bornstein (1996)	6.5	Subset of the Gold Standard Battery (Seashore Rhythm Test, Controlled Oral Word Association Test, Logical Memory Test) - Validation Sample	Researcher selected battery	84 MS	45.2% Chronic Progressive, 22.6% Exacerbated RR, 25.0% Stable RR, 7.1% Unknown	75.0% Female, 25% Male	Not Given	Not Given	Not Given	Not Given	1.00	0.80	0.88
		Subset of the Gold Standard Battery (Seashore Rhythm Test, Controlled Oral Word Association Test, Logical Memory Test) - Cross-Validation Sample	Researcher selected battery								1.00	0.82	0.91

Dent and Lincoln (2000)	9.0	Brief Repeatable Battery of Neuropsychological Tests (BRB-N)	Weschler Memory Scales - Revised	61 MS	42.6% RR, 41.0% SP, 4.9% PP, 6.6% Benign	Not Given	Not Given	42.7 (8.9)	Not Given	Not Given	0.93	0.18	Not given
		10/36 Spatial Recall Test and Symbol Digit Modalities Test from BRB-N	Weschler Memory Scales - Revised								0.93	0.48	Not given
Wilken, Kane, Sullivan, Wallin, Usiskin, Quig, Simsarian, Saunders, Crayton, Mandler, Kerr, Reeves, Fuchs, Manning and Keller (2003)	9.0	ANAM	Researcher selected battery	50 MS	100% RR	70.0% Female, 30.0% Male	72.0% Caucasian, 20.0% African American, 2.0% Hispanic, 2.0% Persian, 4% Other	43.0 (10.9)	15.3 (2.1)	Not Given	Not Given	Not Given	Not given
Negreiros, Mattos, Landeira-Fernandez, Paes and Alvarenga (2008)	5.5	Controlled Oral Word Association Test, Rey Auditory Verbal Learning Test, Enhanced Cued Recall Test	Researcher selected battery	54 MS and 54 HC	100% RR	61.1% Female, 38.9% Male	Not Given	38.7 (1.4)	11.7 (0.5)	Not Given	0.81	0.97	Not given
Lechner-Scott, Kerr, Spencer, Agland, Lydon and Schofield (2010)	10.0	ARCS	Researcher selected battery	45 MS	64.0% RR, 20.5% SP, 7.0% PP, 2.0% PR, 6.0% CIS	83.0% Female, 17% Male	Not Given	46.8 (13.9)	Not Given	12.5 (9.3)	0.86	0.71	0.85
		PASAT	Researcher selected battery								0.68	0.71	0.79

Claesson, Ytterberg, Johansson, Almkvist and Van Koch (2007)	7.0	FRRT - Recall	FRRT, SDMT, PASAT and MMSE	227 MS	59.0% RR, 37.0% SP, 4.0% PP	Not Given	Not Given	Median = 47yrs	Not Given	Median = 11 yrs	0.90	0.25	Not given
		FRRT - Recognition	FRRT, SDMT, PASAT and MMSE								0.70	0.51	Not given
Parmenter, Weinstock-Guttman, Garg, Munschauer and Benedict (2007)	9.0	SDMT	MACFIMS	100 MS and 50 HC	70.0% RR, 30.0% SP, 0.0% PP	78.0% Female, 22.0% Male	94.0% Caucasian, 6.0% African-American	44.6 (8.4)	14.4 (2.0)	Not Given	0.82	0.60	0.84
Scherer, Penner, Rohr et al. (2007)	8.0	FST	Digit Symbol Substitution Test, PASAT	108 MS and 33 HC	74.1% RR, 22.2% SP, 3.7% PP	67.6% Female, 22.4%Male	Not Given	40.6 (8.2)	Not Given	9.5 (6.4)	0.84	0.85	0.91
Benedict, Cox,Thompson, Foley, Weinstock-Guttman and Munscauer (2004)	9.0	MSNQ-P	Researcher selected battery	85 MS and 40 HC	80.0% RR, 20.0% SP/PP	80.0% Female, 20.0% Male	94.0% Caucasian, 6.0% Other	Not Given	14.8 (2.3)	Not Given	0.68	0.80	Not given
		MSNQ-I	Researcher selected battery								0.87	0.84	Not given

Akbar, Honarmand, Kou, Levine, Rector and Feinstein (2010)	9.5	Internet MSNQ-P	BRB-N	82 MS	62.0% RR, 21.0% SP, 7%PP	78.0% Female, 22.0% Male	Not Given	Not Given	15.0 (2.0)	9.5 (7.4)	Not given	Not given	0.57
		Internet MSNQ-I	BRB-N								0.72	0.60	0.71
Vanotti, Benedict, Acion, Caceres and VANEM Workgroup (2009)	8.5	Spanish MSNQ-P	BRB-N	125 MS and 36 HC	86.4% RR, 9.6% SP, 3.2% PP,0.8 PR	67.2% Female, 32.8% Male	Not Given	42.3 (10.5)	Not Given	8.8 (7.0)	Not given	Not given	0.62
		Spanish MSNQ-I	BRB-N								0.91	0.80	0.92
O'Brien, Gaudino-Goering, Shawaryn, Komaroff, Moore and DeLuca (2007)	10.0	MSNQ-P	Researcher selected battery	48 MS and 40 HC	68.8% RR, 21.2% SP, 10.0% PP	80% Female, 20% Male	Not Given	45.1 (9.1)	14.7 (2.1)	14.6	0.52	0.70	0.62
		MSNQ-I	Researcher selected battery								0.66	0.77	0.74

Key:
MSNQ-P = Multiple Sclerosis Neuropsychological Screening Questionnaire – Patient Version
MSNQ-I = Multiple Sclerosis Neuropsychological Screening Questionnaire – Informant Version
SDMT = Symbol Digit Modalities Test
FST = Faces Symbol Test
PASAT = Paced Auditory Serial Addition Test
ANAM = Automated Neuropsychological Assessment Metrics
FRRT = Free Recall and Recognition Test
ARCS = Audio Recorded Cognitive Screen

PP = Primary-Progressive MS
SP = Secondary-Progressive MS
RP = Relapsing-Progressive MS
RR = Relapsing-Remitting MS

MS= Number of participants with MS
HC= Number of healthy control participants

NB: Gender proportions are given for the MS sufferer section of the sample in studies where control groups were used.

An early attempt to produce and validate a screening test for use with Multiple Sclerosis patients was outlined by Beatty et al. (1995). They generated a short battery of neuropsychological tests which they called the Screening Examination for Cognitive Impairment (SEFCI). The SEFCI consisted of: a memory task involving learning a short list of words followed by a delayed recall test; the Shipley Institute of Living Scale (SILS; a 40-item questionnaire test of vocabulary and a 20-item test of verbal abstraction, which is self-completed in around 10 minutes); and the Symbol Digit Modalities Test (SDMT). In the SDMT, patients are presented with a set of abstract symbols that are each paired with one of the numbers 1-9. The patient is asked to provide the corresponding number for a set of symbols, responding either in writing or verbally. The orally administered test is therefore suitable for use with MS sufferers who may experience motor problems. The SEFCI, and a 2-hour, 'gold standard' neuropsychological battery, covering areas of impairment typically associated with MS, were administered to 103 MS patients with a clinically definite diagnosis and 32 matched, healthy controls. Beatty et al. (1995) defined cognitive impairment as performance below the 5th percentile on any of the tests, but took the norms for this from the scores of their healthy control sample. An impaired score on any of the three SEFCI tests correctly identified 100% of MS patients with impairment on 3 or more of the 'gold standard' tests. When used to predict impairment on one of the 'gold standard' tests, the SEFCI was found to have sensitivity of 86% and specificity of 90%. Given that the control sample consisted of only 32 participants, it is possible that the norms used were not representative of the wider population. Using norms from the wider population for the SDMT and SILS reduced sensitivity of the SEFCI to 74.1% but specificity increased slightly to 90.9%. ROC analysis was not used to generate these figures since there was not a requirement to establish a cut-off score for the screening

battery. The SEFCI is severely limited, in requiring approximately 30 minutes to administer. This would not fit well into MS clinics and services where the main requirement of a screening test is quick administration to a large caseload of patients. Faster, more efficient screens are therefore needed. Additionally, one of the SEFCI tests had no published norms, so usefulness in a clinical environment would be severely limited. Beatty et al. (1995) did not provide a breakdown of their MS sample by diagnosis subtypes and both the MS and healthy control samples were highly educated (mean years of education = 14.5), which raises questions about representativeness. Also, participants were tested in either their homes or in a research laboratory. Beatty et al. (1995) found significant differences between scores in these settings so it is questionable whether the findings would transfer to a clinical environment.

Basso, Beason-Hazen, Lynn, Rammohan and Bornstein (1996) also attempted to produce a screening battery. The methodology differed from that of Beatty et al. (1995). Their sample consisted of 84 Multiple Sclerosis patients referred for neuropsychological testing who all met criteria for either probable or definite MS diagnosis. No healthy controls were used. The sample was split into two groups, matched to each other, and the same comprehensive battery of tests was carried out with all participants. A subset of the tests from the battery were then chosen and validated as a screening test, meaning that the screening and reference batteries were not administered by those blind to their results. The screening battery consisted of tests of auditory attention, verbal fluency, verbal learning and sensory-perceptual capacity. Screening test choice was based on brevity of administrations and sensitivity to cerebral dysfunction. In both the matched groups, it was found to have 100% sensitivity and specificity was also high (80% and 81.8% in each of the matched groups). However, the screening battery contained no tests of executive function. This was a notable omission

since it is generally accepted in the literature that executive function is impaired by MS disease action (Rao et al, 1991a; Garcia, Plasencia, Benito, Gomez & Marcos, 2009) and knowledge of executive function is important if rehabilitative or other interventions might be being considered after testing. Basso et al. (1996) accepted that this omission made further testing necessary. Many of the tests used in screening and validation batteries are quite obscure and not in general use, so they have limited clinical application due to this lack of availability. A further limitation is inclusion of probable MS diagnosis in the Basso et al. (1996) sample - those without cognitive problems directly attributable to MS may have been included. Basso et al. (1996) also failed to outline the ethnicities in their sample. The screening battery suffers from the same inefficiency and resource intensity problems as the Beatty et al. (1995) SEFCI tests. It requires 35-50 minutes to carry out. Dent and Lincoln (2000) suggest that the Basso et al. (1996) battery is biased towards physical rather than cognitive problems since it predicted levels of physical disability on the EDSS, contrary to the majority of the literature where no significant correlations are usually seen.

Dent and Lincoln (2000) focussed on validating a screening battery for memory problems. They used the Brief Repeatable Battery of Neuropsychological tests (BRB-N; Bever, Grattan, Panitch and Johnson, 1995) and examined a variety of its subtests in relation to the Wechsler Memory Scale – Revised (WMS-R; Wechsler, 1987) as a gold standard. The BRB-N consists of 5 subtests – the SDMT, the 10/36 spatial recall test (10/36 SRT) which tests visual-spatial memory, the Bushke Selective reminding test which requires recall of auditorily presented word lists, the Paced Auditory Serial Addition Test (PASAT; a test of attention) and the Controlled Oral Word Association Test (COWAT; a test of verbal fluency). ROC analysis was carried out for each of these tests to determine optimum cut off points. Dent and Lincoln (2000) found that use of

the SDMT and 10/36 SRT produced a high sensitivity of 93% but a poor specificity of 48%. This means that use of these tests would be likely to produce a good number of false positive results. Other combinations of BRB-N tests also produced low specificities. The study methodology is notable, however, for using an element of random sampling. While other studies have taken all MS sufferers who volunteered to participate, Dent and Lincoln (2000) took a random sample of 62 participants from a total of 92 who volunteered to take part. This should make the sample more representative but a number of factors undermine this in the study. Some participants were described in the paper as having 'probable MS', a term used by neurologists prior to giving a clinically definite diagnosis. Impairments found in the sample could not be definitely attributed to MS disease action. There were higher than typical numbers of participants with a secondary progressive diagnosis and no breakdown of ethnicity or gender of the sample was provided. Since no exclusion criteria were provided for the study, other neurological conditions or a history of alcohol abuse could be co-morbid in the sample. Validating a test to use in MS patients requires a 'purer' sample than the one used by Dent and Lincoln (2000) for the results to be generalisable to clinical practice.

Basso et al. (1996) suggested computerising their screening battery and Wilken et al. (2003) carried out an assessment of a computerised test. They described a number of benefits from using computerised tests such as shorter completion times, ability to generate alternate forms of tests to prevent practice effects and increased accuracy in recording completion times. Wilken et al. (2003) compared an established computerised screening test – the Automated Neuropsychological Assessment Metrics (ANAM) – with a 'gold standard' battery of neuropsychological tests. A total of 50 participants were recruited from MS clinics local to the experimenters in the United States. All

participants had a diagnosis of relapsing-remitting MS and had not received treatment with the usual medications for this (interferon or glatiramer acetate) within the last 12 months. The absence of other MS subtypes in the sample means results cannot be generalised to a typical MS population. The neuropsychological battery was administered before the ANAM but since marking of the ANAM is computerised, this should not introduce any confirmation bias to the results. However, given that the full set of testing lasted two and a half hours and was carried out in a single session, fatigue may have presented a confounding factor in the Wilken et al. (2003) study. Correlations between the ANAM and 'gold standard' measures were calculated and results on subtests of the ANAM were found to have statistically significant correlations to tests of memory, problem solving, verbal fluency, attention and processing speed in the neuropsychological battery. No ROC analysis, however, was carried out; instead, an impairment index was calculated from the findings and compared to indices from other studies. This showed the ANAM to be valid in identifying levels of impairment similar to other tests.

All studies considered so far have used samples from English speaking countries but neuropsychological tools need to be applicable and valid in many cultures and languages. Negreiros, Mattos, Landeira-Fernandez, Paes and Alvarenga (2008) pointed out the importance of cross-cultural work in neuropsychology and the lack of research into validated tools for use with Portuguese-speaking MS patients in Brazil. Consecutively admitted MS patients with Relapsing-Remitting MS (RRMS), and no other clinical conditions which interfere with cognition, were recruited to the study and matched with a control group of relatives and friends of the researching medical team. All participants completed a seven test neuropsychological battery including tests of attention, information processing speed, short- and long-term memory, verbal fluency,

abstract thinking and a dementia screening test. This battery produced a set of 26 variables indicating level of cognitive abilities. All 26 variables demonstrated a statistically significant difference between RRMS and control samples. 51.8% of RRMS patients compared to the control group showed cognitive impairment on four indices or more. Level of impairment did not correlate with physical disability as measured by the EDSS. Using the percentage of the RRMS sample scoring below the 5^{th} percentile as a measure of sensitivity of each variable, the most sensitive tests were grouped into three proposed screening tests and ROC analysis was carried out for each one. Use of three tests – the COWAT verbal fluency test, the Rey Auditory Verbal learning test and the Enhanced Cued Recall test – as a screening tool gave a sensitivity of 80.6% and a specificity of 97.2%. While a ROC curve was plotted in the report no area under the curve or accuracy figure was quoted. The applicability of the findings was limited by the use of a sample which only consisted of RRMS sufferers. Since an RRMS diagnosis only applies to around half of MS patients, the study would need replicating in a sample more representative of the general MS population. Additionally, the methodology used did not allow for blinding in terms of marking the screening tool and reference criteria separately. Results were only applicable to a Portuguese speaking, Brazilian population and had limited relevance to an English speaking MS population. Additionally, potential participants with EDSS scores above 6.5 were excluded, meaning the screening test may not be valid for those with further advancement of the physical symptoms of MS. The Negreiros et al. (2008) study cannot, therefore, be considered a strong validation study since the screening tool and reference criteria were not specified in advance.

Use of a short battery of tests such as the Brief Repeatable Battery or the Negreiros et al. (2008) battery allows assessment of a broad range of cognitive domains

affected by MS disease action. However, they require a large amount of time to administer and usually require trained practitioners for interpretation. Lechner-Scott et al. (2010) therefore investigated a screening battery which could be administered to unsupervised MS patients using an audio device. This administration reduces the involvement of trained professionals to only the 5 minutes required for test scoring. The Audio Recorded Cognitive Screen (ARCS) had previously been validated in comparison to other cognitive tests and normative data was derived using a community sample with no history of impairments (Schofield et al., 2010). Lechner-Scott et al. (2010) administered the ARCS to a sample of 127 MS patients, of which 45 participants also went on to complete a 'gold standard' neuropsychology battery and 87 completed the PASAT – a test of auditory information processing speed and flexibility. ROC analysis indicated that at equivalent specificity, the ARCS had superior sensitivity of 86% compared to 68% for the PASAT. Area under the ROC curve (AUROC) was also superior for the ARCS compared to PASAT (0.85 versus 0.79). Despite these findings, the ARCS does take 34 minutes for a patient to complete which, although not practitioner time, may still be longer than MS patients would be willing to spare on top of usual review appointments at MS clinics. The requirement of a separate testing room may also be problematic for clinics with limited facilities and large caseloads. The ARCS allows calculation of cognitive domain scores and these correlated significantly with equivalent scores from the gold standard battery in memory and executive function but not for the visuospatial domain. Despite its lengthier administration time, the ARCS may be a useful short form battery which could be used in place of fuller neuropsychological testing if required. The Lechner-Scott et al. (2010) study also benefitted from using separate testers for the ARCS and gold standard testing. While a range of clinical subtypes were used, details of participant ethnicities were not given

and the sample also included those with a diagnosis of Clinically Isolated Syndrome[1] which does not constitute a confirmed MS diagnosis. The sample used could not be considered representative of the general MS population. The Lechner-Scott et al. (2010) article does not provide details of exclusion criteria so co-morbid conditions may also have confounded the data.

If a quick test in a single domain of functioning such as memory or executive function could be found to have adequate sensitivity, specificity and accuracy, then this would be preferable to use instead of short batteries. Claesson, Ytterberg, Johansson, Almkvist and Von Koch (2007) used a single memory test as a screening instrument for cognitive impairment in MS and found it to have good sensitivity but poor specificity. They chose the Free Recall and Recognition Test (FRRT) since recent memory may be representative of wider cognitive dysfunction in MS. In the FRRT, patients are visually and audibly presented with a stimulus list of twelve concrete nouns. After withdrawal of the list, patients are given two minutes for free recall of the words presented. A recognition trial is also carried out using an equal number of the words and distractor items. A serious weakness of the methodology was that Claesson et al. (2007) only used another single test as a gold standard – the SDMT - which tests processing speed and not memory. The FRRT recall score was found to have 90% sensitivity and 25% specificity at optimum cut off score. The FRRT recognition optimum cut off score showed a poor sensitivity of 70% and low specificity of 51%. These values can only be considered to represent the ability of the FRRT to identify those with slowed cognitive processing due to the narrow scope of the reference criterion. In comparison to other tests – the PASAT, SDMT and Mini Mental State Examination (MMSE, a short

[1] Clinically Isolated Syndrome is a first neurological event caused by either inflammation or demyelination in at least one site of the Central Nervous System and lasting more than twenty four hours. In order to meet diagnostic criteria for MS it is necessary that more than one such event has occurred for the patient. A person with a Clinically Isolated Syndrome may or may not go on to develop MS.

cognitive screen widely used in neurology settings) – the FRRT was better at distinguishing between MS patients when classed as either mild, moderate or severely disabled on the EDSS. However, this would cast further doubt on the FRRT's value as a screening tool as there is much literature suggestive that EDSS scores do not correlate with level of cognitive impairment in MS.

Parmenter, Weinstock-Guttmann, Garg, Munschauer and Benedict (2007) investigated the possibility of using the SDMT as a screening test in itself. They administered the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS; Benedict et al., 2002) battery of neuropsychological tests, which includes the SDMT, to a sample of 100 MS patients and 50 controls. They separated out the SDMT scores and compared them to the results of the rest of the MACFIMS as the 'gold standard' measure. The MACFIMS is a battery of neuropsychological tests established by a consensus committee of psychologists and neuropsychologists from the USA, Canada, UK and Australia (Benedict et al., 2002). Parmenter et al. (2007) also administered the Beck Depression Inventory – Fast Screen (BDI-FS) to account for the influence of mood on the test scores. The BDI-FS is frequently used in place of the longer BDI with MS patients since it omits some questions which MS patients would score highly on due to their symptoms rather than depression. The researchers found statistically significant correlations ($p<0.001$) between the SDMT and all the tests in the reduced version of the MACFIMS and this effect persisted even when SDMT results were controlled for the effect of depression. MS patients' scores on the SDMT were significantly lower than controls ($p<0.001$). At the optimum cut off point for distinguishing impaired from non-impaired patients, the SDMT was found to have a sensitivity of 0.82 and specificity of 0.60. A ROC curve was plotted and the area under the curve was 0.84, suggesting high accuracy. The researchers suggest further

investigations should examine ways of using the SDMT and another screening test – the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) - together. The study did not use blinding for administration of reference and screening tests, although SDMT scores were interpreted separately from the other MACFIMS tests. The study methodology was clearly described and all relevant demographic details were provided for the sample used, which benefitted from consisting of those with a clinically definite MS diagnosis and has exclusion criteria ensuring other confounding conditions were not included.

The majority of literature on screening for cognitive impairments in MS uses established tests. However, the final two screening tools considered in the current paper are both novel. The Faces Symbol Test (FST) was developed and validated in a multi-centre trial in Berlin, Germany by Scherer et al. (2007). The FST is based upon the Digit Symbol Substitution Test (DSST) in the Wechsler Adult Intelligence Scale – Revised (WAIS-R; Wechsler, 1981) which is still present in the latest version of the Wechsler scales as the Coding subtest (WAIS-IV, Wechsler, 2008). The FST uses pictures of faces as the stimulus, underneath which the corresponding abstract symbol from the key must be written. It is a parallel version of the DSST but uses faces instead of numbers, making it culture and language neutral. Scherer et al. (2007) validated their test by administering it to a sample of 108 MS outpatients using the parallel version of the DSST along with the PASAT as a reference criterion. The FST was found to have sensitivity of 0.84 and a specificity of 0.85. When a ROC curve was plotted, the area underneath the curve was found to be 0.91. The FST, therefore, has high levels of accuracy, sensitivity and specificity. However, these results were obtained using only two tests as a gold standard – one of which was a parallel version of the one under investigation. Such good results are not, therefore, surprising given this methodological

weakness. Sensitivity, specificity and accuracy were lower when a neurologist's assessment of cognitive status was used as reference standard, thus removing the DSST/FST confound. Both the PASAT and DSST are tests of attention and processing speed, so the FST cannot claim to be adequately predictive of impairments of memory – the most commonly experienced cognitive dysfunctions in MS. Scherer et al.'s (2007) paper was written in a very transparent way and was explicit about the research protocol used. A great strength of his research was the use of blinding by having separate assessors carry out the FST and the PASAT tests. A shortcoming of the FST, however, is the demand it makes on MS patient's motor abilities. Scores on the FST were significantly correlated to participants' EDSS scores, suggesting physical disability is a confounding factor in the results. Scherer et al. (2007) recommended the SDMT or PASAT be used instead of the FST with MS patients who have motor difficulties. Additionally, the paper did not give a breakdown of the ethnic background of the sample used, which raises doubts about the generalisability of the findings for a test that seeks to have global applicability.

The Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) has already been mentioned above as a novel tool specifically developed to screen for cognitive dysfunction in MS. Benedict et al. (2004) sought to validate it using a sample of 85 participants taken from MS clinics across the United States. The development of the MSNQ was previously described by Benedict et al. (2003) where a set of 15 statements about possible cognitive problems was derived from an initial set of 80 using Rasch analysis. These statements were provided in questionnaire form to both patients and informants who knew the patient well. The MSNQ has good internal consistency with Cronbach's alpha of 0.93 and 0.94 for the patient- and informant-administered versions respectively. Benedict et al. (2004) therefore sought to validate the two versions of the

questionnaire in a large sample by comparison to a battery of neuropsychological tests equivalent to the MACFIMS. The Fast Screen version of the Beck Depression Inventory (BDI-FS) was also administered. The reference criterion was administered following completion of the MSNQ, so screening test scores were not influenced by reference criterion scores. The MSNQ-Patient version was found to have sensitivity of 0.80 and specificity of 0.68. However, this result was only found when using the MSNQ-Patient to identify participants who were affected by either depression or cognitive impairment. The questionnaire correctly identified fewer participants when cognitive impairment alone was the criterion. A score of above 4 on the BDI-FS was used as the criterion for being affected by depression which is low, considering a score of 4 signifies only mild depression (Beck, Steer & Brown, 2000). In comparison, the informant version of the questionnaire had a sensitivity of 0.87 and a specificity of 0.84 and was correlated with cognitive dysfunction rather than depression. While clear exclusion criteria were used to avoid confounds from other diagnoses that may cause cognitive impairments, the sample of MS patients used was 80% female and 94% Caucasian. Even when considering higher prevalence levels of the diagnosis in such groups, representativeness of the sample was not ideal. However, Benedict et al. (2004) were clear about the limitations of the MSNQ and their study.

Exclusion criteria for the Benedict et al. (2004) study precluded those with motor impairments from taking part. The MSNQ and FST are both 'paper and pencil' tests but impairment and degeneration of motor and sensory abilities are notable symptoms of MS. For some sufferers, these tests may not be possible to administer. For the MSNQ, where the patient version is less valid as a measure of cognitive impairment, this is all the more important if MS patients have no reliable informant.

Akbar et al. (2010) tested an adapted version of the MSNQ which allows the questionnaire to be administered over the internet. Technological adaptations could be used to enable MS patients with motor difficulties to interact with this version of the test. Akbar et al.'s (2010) ROC analysis was carried out by comparing these computerised informant and patient versions of the MSNQ administered over the internet with the Brief Repeatable Battery (BRB) as a gold standard. They also examined correlations between the BRB tests, a measure of depression and the MSNQ versions. The patient version only correlated with two of the BRB tests and also with the depression measure. The informant version correlated with all five BRB tests but also, contrary to Benedict et al.'s (2004) findings, with the depression measure. There were strong findings for test-retest and internal reliability. The informant-MSNQ had the superior sensitivity of 0.72 and specificity of 0.60. Accuracy of the tests as indicated by the area under the curve was 0.707 for the informant version and 0.566 for the patient version. These findings suggest that internet administration of the informant-MSNQ makes it less valid as a measure of cognitive impairment and more likely to be confounded by low mood in the MS patient. Caution should be exercised in comparing these findings with those of Benedict et al. (2004) since different 'gold standard' criteria were used. Akbar et al. (2010) also highlighted difficulties with internet administration of the screening tool since some informant participants took over 45 minutes to complete the 15 questions and were excluded from the study. While the experimenters could not state why participants appeared to take breaks in completing the questionnaire they excluded these participants on the basis they had not been properly engaged with the research process.

Vanotti, Benedict, Acion, Caceres and the VANEM Workgroup (2009) validated a version of the MSNQ translated into Spanish for use in Argentina. They

wished to test the MSNQ's psychometric properties in a non-English speaking population. They used the BRB-N as their gold standard, as this has also been previously translated into Spanish and used a sample of 125 MS patients plus 36 healthy controls. All participants received both the screening test and the reference criterion. As with the internet version, the informant-MSNQ correlated significantly with all psychometric tests in the gold standard battery. The patient version, however, only correlated significantly with the Selective Reminding Test (SRT). The Spanish translation was found to have good internal consistency with Cronbach's alpha of more than 0.90 whenever implemented except for with healthy controls. Test-retest reliability analysis produced an interclass correlation of 0.95 for the Patient-MSNQ and 0.94 for the Informant-MSNQ. Using ROC analysis to test for validity gave poor results for the Patient-MSNQ. While Vanotti et al. (2009) did not report sensitivity or specificity values from this analysis, the Area under the Curve (AUC) was only 0.62. In comparison, the Informant-MSNQ was found to have a high AUC of 0.92, sensitivity of 0.91 and specificity of 0.80, showing it to be a highly valid measure of cognitive impairment in a Spanish population and strengthening the case for its wider use. The direct applicability of these findings in an English-speaking population is low but the study by Vanotti et al. (2009) strengthens the case for clinical use of the MSNQ in other countries. However, the methodology outlined no procedures for blinding in the administration of the screening or reference tests, which is of importance since all testing was conducted in one session by the same experimenters. Knowledge of previous scores may have influenced scoring in the reference standard testing or created demand effects in participants.

O'Brien et al. (2007) also examined the MSNQ and confirmed the findings of Benedict et al. (2004). They administered both versions of the MSNQ to a sample of 48

MS patients and 40 controls and plotted a ROC curve by comparing to a gold standard battery of neuropsychological tests that examined pre-morbid function, attention, processing speed, verbal and working memory, verbal learning, executive function and visual-spatial abilities. Results were similar to previous studies, with the Patient-MSNQ only correlating significantly with two tests from the neuropsychological battery, while the Informant version significantly correlated to a much wider range of tests. Correlations previously found with the Beck Depression Inventory were also replicated. Unlike previous studies, O'Brien et al. (2007) also investigated the relationship of the MSNQ to measures of everyday functioning. The Patient-MSNQ significantly correlated to self-report measures of daily function but the Informant-MSNQ did not. ROC analysis showed the Informant-MSNQ to be more accurate than the Patient version but O'Brien et al. (2007) concluded that neither measure was adequate to discriminate between impaired and non-impaired groups. Both measures were found to have optimal sensitivity and specificity at much lower values than found by Benedict et al. (2004). Due to its strong correlation to other measures of cognitive function, they recommended the use of the Informant-MSNQ as a screening tool for identifying patients who may benefit from further neuropsychological testing, but did not recommend it be used in place of such extensive tests due to its poor accuracy. They suggested further research is needed to establish the quality of the MSNQ as a screening instrument. The study methodology was reported by O'Brien et al. (2007) very comprehensively and tables of all statistical test results and ROC curves were provided, making this a very transparent and replicable study. It is important to note that, while the sample used was broadly representative of an MS population, it was not as large as in other studies (n=48, see Table 1) and also extensive neuropsychological testing was

used without any measure of fatigue applied in the test battery in order to ascertain the impact fatigue had on the participants results.

Discussion

A total of 11 screening tools have been examined by the 13 studies covered in the current review. These tools have varied from short batteries through to single tests. Duration to complete the single tests has been between 90 seconds for the Symbol Digit Modalities Test (SDMT) through to around half an hour for the Audio Recorded Cognitive Screen (ARCS). The Screening Examination for Cognitive Impairment (SEFCI; Beatty et al., 1995) and the short battery used by Basso et al. (1996) have both been found to have good sensitivity and specificity, while the BRB-N (Dent & Lincoln, 2000) had poor specificity but good sensitivity. However, these short batteries can be still criticised for being too long in terms of completion times. For the single tests, performance is variable, although the ARCS (Lechner-Scott et al., 2010), SDMT (Parmenter et al., 2007) and Faces Symbol Test (FST; Scherer et al., 2007) demonstrated very good levels of sensitivity, specificity and accuracy. Validation of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) has provided results that suggest it has good sensitivity and specificity, although there is some variation in these values. A reliable finding is that the informant version of the questionnaire is a better predictor of cognitive impairments than the patient version. The MSNQ was found to perform best when translated into Spanish rather than in English (Vanotti et al., 2009) although 'gold standards' used were not comparable across studies.

Sampling methods used in the reviewed studies have made them broadly representative of the population of MS sufferers; the only exception to this was the occurrence of MS samples often having high levels of education compared to the general population. Although the majority of studies were carried out with samples from the USA, some tests have been able to be validated in other countries such as England, Germany, Brazil and Argentina. The most widely tested tools in the current

review were the SDMT and the MSNQ with the latter of these particularly being repeatedly investigated. It is possible that there was bias in the sampling used for the current review as only studies published in English were included. It was hoped that a study published in French would be able to be included but while the abstract had been translated, unfortunately the article itself had not.

A further criticism of sampling methods is the frequent omission in many studies of participants with diagnoses other than Relapsing-Remitting MS. Little consideration was given in any of the studies to whether screening tests were more or less effective with different diagnoses or whether particular diagnoses were associated with differing degrees or patterns of impairments. This could be a productive line of enquiry for further research given that Wachowius, Talley, Silver, Heinze and Sailer (2005) have found differing patterns and higher rates of impairment among patients with Primary-Progressive MS compared to those with Secondary-Progressive MS. Additionally, it was frequent for participants with high Expanded Disability Status Scale (EDSS) scores to be excluded from studies, meaning many findings cannot be generalised to the more physically disabled of the MS population.

The reference criteria used for each study were different from each other, with some studies using established neuropsychological test batteries and others devising their own based on knowledge of the literature. An adequate validation of a screening tool or battery should use established and widely accessible tests. In two studies, the reference criteria were limited to only one or two other tests that were not comprehensive in the range of cognitive domains they considered. Later validations were stronger in this aspect of methodology by using batteries that have established research consensus behind them, such as the Brief Repeatable Battery of Neuropsychological tests (BRB-N) or the Minimal Assessment of Cognitive Function

in Multiple Sclerosis (MACFIMS). The variation in gold standard batteries in earlier studies in the current review means their findings could not be compared directly to each other.

It is notable that few of the studies in the current review described using any measures to blind experimenters from participants' previous scores. For some screening tools, high scores clearly would be expected to associate with cognitive impairment and if the same researcher carries out the reference battery, they may err on the side of harshness in scoring later tests if there is ambiguity in participant responses. A methodological strength of many studies in this sense was that the screening battery was at least carried out prior to the gold standard.

In the studies considered by the present review, the medication which participants were prescribed was not explicitly investigated as part of their research hypotheses. However, in a number of studies, certain medications were included as an exclusion criterion. Medication status is an important variable for future research to consider with regard to testing for cognitive impairment given the degree to which it is used as a part of treatment of the condition.

A variety of screening tests for cognitive impairment exist. Screening tools investigated in early research were simply small groups of tests that were subsets of larger batteries selected for their ability to distinguish impaired from non-impaired participants effectively. Later research has attempted to assess single, short screening tools for sensitivity, specificity and accuracy using comparison to a gold standard reference criterion. A variety of such gold standards have been used but these are typically a battery of tests investigating the domains of cognitive function affected by MS – processing speed, memory, attention, working memory, executive function and visuospatial perception. A variety of short screening tools are available that mostly

investigate either processing speed or memory as a proxy for the wider range of impairments found in MS. Based on quoted levels of accuracy in the literature the best tests to use are the Faces Symbol Test (FST; Scherer et al., 2007) and the Symbol Digit Modalities Test (SDMT; Parmenter et al., 2007). The Spanish translation of the Multiple Sclerosis Neuropsychological Questionnaire also performs well in this regard although clearly this would not be recommended for use in English-speaking countries. However, using the assessment tool provided by the Public Health Resource Unit (2006) the best quality studies were those of O'Brien et al. (2007) and Lechner-Scott et al. (2010) so their findings can be considered the most informative about the true state of the tests they investigated. This would suggest that further investigation of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) would be of value since O'Brien et al. (2007) found it to have a poor sensitivity, specificity and accuracy in both of its forms contrary to the findings of the Benedict et al. (2004) study which still has a strong quality score. Clarification of the usefulness of the MSNQ as a screening tool is clearly needed. Lechner-Scott et al. (2010) found strong sensitivity, specificity and accuracy values for the Audio Recorded Cognitive Screen in their study suggesting it is worthy of further consideration for use with MS patients. It is also notable that Parmenter et al. (2007)'s investigation of the SDMT as a screening tool also scored well on the quality assessment tool further strengthening the case for its' use clinically. The preferred screening test chosen by a clinic is likely to be determined by the situation of both the client and the service assessing them. Given that fatigue is frequently problematic in the MS population (Barak & Achiron, 2006) and that MS clinics frequently have large caseloads, with resultant demands on clinic time for each patient, the optimal choice of screening tool is strongly dictated by the administration time for each test.

Given the quality of the research paper and the strong values of sensitivity, specificity and accuracy for the SDMT this would be the tool which the current review would recommend for use in clinics as the best way to screen for cognitive deficits in MS.

References

Akbar, N., Honarmand, K., Kou, N., Levine, B., Rector, N. & Feinstein, A. (2010) Validity of an internet version of the multiple sclerosis neuropsychological questionnaire. *Multiple Sclerosis*, 16 (2), 1500-1506.

Barak, Y. and Achiron, A. (2006) Cognitive fatigue in multiple sclerosis: Findings from a two-wave screening project. *Journal of the Neurological Sciences*, 245, 73-76.

Basso, M.R., Beason-Hazen, S., Lynn, J., Rammohan, K. and Bornstein, R.A. (1996) Screening for cognitive dysfunction in multiple sclerosis. *Archives of Neurology*, 53, 980-984.

Beatty, W.W., Paul, R.H., Wilbanks, S.L., Hames, K.A., Blanco, C.R. and Goodkin, D.E. (1995) Identifying multiple sclerosis patients with mild or global cognitive impairment using the screening examination for cognitive impairment (SEFCI). *Neurology*, 45, 718-723.

Beck, A.T., Steer, R.A., Brown, G.K. (2000) *BDI-Fast Screen for Medical Patients: Manual*, San Antionio, TX: Psychological Corporation.

Benedict, R.H.B., Fischer, J.S., Archibald, C.J., Arnett, P.A., Beatty, W.W., Bobholz, J., Chelune, G.J., Fisk, J.D., Langdon, D.W., Caruso, L., Foley, F., LaRocca, N.G., Vowels, L., Weinstein, A., DeLuca, J., Rao, S.M. and Munschauer, F. (2002) Minimal neuropsychological assessment of MS patients: A consensus approach. *The Clinical Neuropsychologist*, 16 (3), 381-397.

Benedict, R.H.B., Munschauer, F., Linn, R., Miller, C., Murphy, E., Foley, F., & Jacobs, L. (2003) Screening for Multiple Sclerosis Cognitive Impairment Using a Self-Administered 15-Item Questionnaire. *Multiple Sclerosis*, 9, 95-101.

Benedict, R.H.B., Cox, D., Thompson, L.L., Foley, F., Weinstock-Guttmann, B. & Munschauer, F. (2004) Reliable Screening for Neuropsychological Impairment in Multiple Sclerosis. *Multiple Sclerosis*, 10, 675-678.

Bever, C.T., Grattan, L., Panitch, H.S. and Johnson, K.P. (1995) The brief repeatable battery of neuropsychological tests for MS: A preliminary study. *Multiple Sclerosis*, 1, 165-169.

Claesson, I.M., Ytterberg, C., Johansson, S., Almkvist, O. and Von Koch, L. (2007) Rapid cognitive screening in multiple sclerosis accomplished by the Free Recall and Recognition Test. *Multiple Sclerosis*, 13, 272-274.

Delis, D.C., Kramer, J.H., Kaplan, E. & Ober, B.A. (2000) *California Verbal Learning Test Manual Second Edition: Adult Version*, San Antionio, TX: Psychological Corporation.

Dent, A. and Lincoln, N.B. (2000) Brief report: Screening for memory problems in multiple sclerosis. *British Journal of Clinical Psychology*, 39, 311-315.

Engel, C., Greim, B. and Zettl, U.K. (2007) Diagnostics of cognitive dysfunctions in multiple sclerosis. *Journal of Neurology*, 254 (Suppl. 2), II/30-II/34.

Garcia, M.C., Plasencia, P.M., Benito, Y.A., Gomez, J.J.B. & Marcos, A.R. (2009) Executive function and memory in patients with relapsing-remitting multiple sclerosis. *Psicothema*, 21 (3), 416-420.

Gronwall, D.M. (1977) Paced Auditory Serial Addition Task: A Measure of Recovery from Concussion. *Perceptual and Motor Skills*, 44(2), 367-373.

Hirst, C., Ingram, G., Pickersgill, T., Swingler, R., Compston, D.A.S. and Roberston, N.P. (2009) Increasing prevalence and incidence of multiple sclerosis in south east Wales. *Journal of Neurology, Neurosurgery and Psychiatry*, 80, 386-391.

Jaesche R., Guyatt G.H., & Sackett D.L. (1994) Users' guides to the medical literature: VI. How to use an article about a diagnostic test. *Journal of the American Medical Association*, 271 (5), 389-391.

Kurtzke, J.F. (1983) Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). *Neurology*, 33, 1444-1452.

Lechner-Scott, J., Kerr, T., Spencer, B., Agland, S., Lydon, A. and Schofield, P.W. (2010) The Audio Recorded Cognitive Screen (ARCS) in patients with multiple sclerosis: a practical tool for multiple sclerosis clinics. *Multiple Sclerosis*, 16 (9), 1126-1133.

Negreiros, M.O., Mattos, P., Landeira-Fernandez, J., Paes, R.A. and Alvarenga, R.P. (2008) A brief neuropsychological screening test battery for cognitive dysfunction in Brazilian multiple sclerosis patients. *Brain Injury*, 22 (5), 419-426.

National Institute for Health and Clinical Excellence (2003) *Multiple Sclerosis: Mangement of Multiple Sclerosis in Primary and Secondary Care*, London, UK: National Institute for Clinical Excellence.

O'Brien, A., Gaudino-Goering, E., Shawaryn, M., Komaroff, E., Moore, N.B. & DeLuca, J. (2007) Relationship of the multiple sclerosis neuropsychological

questionnaire (MSNQ) to functional, emotional and neuropsychological outcomes. *Archives of Clinical Neuropsychology*, 22, 933-948.

Parmenter, B.A., Weinstock-Guttmann, B., Garg, N., Munschauer, F. and Benedict, R.H.B. (2007) Screening for Cognitive Impairment in Multiple Sclerosis Using the Symbol Digit Modalities Test. *Multiple Sclerosis*, 13, 52-57.

Public Health Resource Unit (2006) *Critical Appraisal Skills Programme (CASP): 12 Questions To Help You Make Sense Of A Diagnostic Test Study*, Oxford, UK: Public Health Resource Unit.

Rao, S.M., Leo, G.J., Bernardin, L. and Unverzagt, F. (1991a). Cognitive dysfunction in multiple sclerosis: I. Frequency, patterns and prediction. *Neurology*, 41, 685-691.

Rao, S.M., Leo, G.J., Ellington, L., Nauertz, T., Bernardin, L. and Unverzagt, F. (1991b). Cognitive dysfunction in multiple sclerosis: II. Impact on employment and social functioning. *Neurology*, 41, 692-696.

Rosati, G. (2001) The prevalence of multiple sclerosis in the world: An update. *Neurological Sciences*, 22, 117-139.

Scherer, P. (2007) Cognitive screening in multiple sclerosis. *Journal of Neurology*, 254 (Suppl. 2), II/26-II/29.

Scherer, P., Penner, I.K., Rohr, A., Boldt, H., Ringel, I., Wilke-Burger, H., Burger-Deinerth, E., Isakowitsch, K., Zimmermann, M., Zahrnt, S., Hauser, R., Hilbert, K., Tiel-Wilck, K., Anvari, K., Behringer, A., Peglau, I., Friedrich, H., Plenio, A., Benesch, G., Ehret, R., Nippert, I., Finke, G., Schulz, I., Bergtholdt, B., Breitkopf, S., Kaskel, P., Reischies, F. & Kugler, J. (2007) The Faces Symbol Test, a Newly Developed

Screening Instrument to Assess Cognitive Decline Related to Multiple Sclerosis: First Results of the Berlin Multi-Centre FST Validation Study. *Multiple Sclerosis*, 13, 402-411.

Schofield, P.W., Lee, S.J., Lewin, T.J., Lyall, G., Moyle, J., Attia, J. and McEvoy, M. (2010) The Audo Recorded Cognitive Screen (ARCS): A flexible hybrid cognitive test instrument. *Journal of Neurology, Neurosurgery and Psychiatry*, 81 (6), 602-607.

Sellner, J., Kraus, J., Awad, A., Milo, R., Hemmer, B. and Stuve, O. (2011) The increasing incidence and prevalence of female multiple sclerosis – a critical analysis of potential environmental factors. *Autoimmunity Reviews*, 10, 495-502.

Vanotti, S., Benedict, R.H.B., Acion, L., Caceres, F. & VANEM Workgroup (2009) Validation of the multiple sclerosis neuropsychological questionnaire in Argentina. *Multiple Sclerosis,* 15, 244-250.

Wachowius, U., Talley, M., Silver, N., Heinze, H.J. and Sailer, M. (2005) Cognitive Impairment in Primary and Secondary Progressive Multiple Sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 27, 65-77.

Wechsler, D. (1981) *Manual for the Weschler Adult Intelligence Scale – Revised.* New York, NY: The Psychological Corporation.

Wechsler, D. (1987) *Weschler Memory Scale – Revised Manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008) *Weschler Adult Intelligence Scale –Fourth Edition.* San Antonio, TX: Pearson.

Wilken, J.A., Kane, R., Sullivan, C.L., Wallin, M., Usiskin, J.B., Quig, M.E., Simsarian, J., Saunders, C., Crayton, H., Mandler, R., Kerr, D., Reeves, D., Fuchs, K., Manning, C. and Keller, M. (2003) The utility of computerized neuropsychological assessment of cognitive dysfunction in patients with relapsing-remitting multiple sclerosis. *Multiple Sclerosis*, 9, 119-127.

Research Report

Validation of Combined Use of the Multiple Sclerosis Neuropsychological Questionnaire and the Symbol Digit Modalities Test to Screen for Cognitive Dysfunction in Multiple Sclerosis

Abstract

Multiple Sclerosis is an autoimmune disease thought to be triggered by an infection which sets off a genetically determined immune dysfunction (Pugliatti, Sotgiu and Rosati, 2002). Disease action causes depletion of insulating myelin around nerve fibres, damage to axons and inflammation around nerve cells. Cognitive domains affected include processing speed, memory, attention, executive function and visuospatial perception (Rao et al., 1991a; Julian, 2011). Short screening tests have been developed to identify sufferers with these impairments. Parmenter et al. (2007) suggest joint investigation of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) and the Symbol Digit Modalities Test (SDMT). Morrow et al. (2010) have demonstrated their reliability and the current study aimed to establish validity in comparison to a gold standard battery of tests.

A sample of 112 participants with a definite diagnosis of Multiple Sclerosis were recruited through Multiple Sclerosis clinics held at a local hospital and drop in surgeries held by MS Specialist nurses at community resource centres. From this initial group, a sub-sample of 26 participants underwent further 'gold standard' testing with the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS; Benedict et al., 2002) battery.

Logistic regression was used to obtain combined scores for the two screening tests. The MSNQ-Patient version combined with the SDMT produced a ROC curve with Area Under the Curve = 0.773 (n=26, p=0.020, std. error = 0.097, 95% confidence interval = 0.583-0.963). At the optimum cut-off score of -0.1275, the combined score had a sensitivity of 0.727 and a specificity of 0.800. ROC analysis on combined scores from the MSNQ-Informant version and SDMT tests produced a ROC curve with AUC = 0.808 (n=20, std. error = 0.103, p = 0.020, 95% confidence interval = 0.605 -1.000). At the optimum cut-off score of -0.7520, the combined score had a sensitivity of 0.889 and specificity of 0.636. Combined use of the MSNQ-Informant and SDMT is more accurate than the use of these tests individually. Use of the MSNQ-Patient and SDMT in combination is more accurate to using the SDMT alone. The MSNQ-Patient is more accurate when using individually than in combination with the SDMT.

Samples were underpowered for the logistic regression and ROC analysis so results should be interpreted tentatively. More comprehensive validation studies are required, although the current analysis indicated that combined utilization of the tests can provide superior results compared to individual use. Options for future research are suggested and discussed.

Introduction

Multiple Sclerosis (MS) is considered an autoimmune disease on the basis that it appears the condition progresses as the body is attacked by its own immune system. The process is thought to be triggered by an infection which sets off a genetically determined immune dysfunction (Pugliatti, Sotgiu & Rosati, 2002). The result is inflammation around nerve cells, damage to axons and depletion of myelin around nerve fibres in the central nervous system. Myelin acts as an insulator around these nerve fibres, so disruption of communication throughout the central nervous system results. MS exhibits 'disease heterogeneity' (Olazaran et al., 2009): symptoms vary from sufferer to sufferer and there is no clear progression or specific set of symptoms.

Epidemiology

A number of factors have been related to developing MS of which distance from the equator is one among many. Simpson, Blizzard, Otahal, Van der Mei and Taylor (2011) suggest that differences in exposure to ultra-violet radiation and resultant lack of production of Vitamin D are the chief factors involved in this process. However, Benito-Leon (2011) suggests that findings across the globe indicate similar prevalence regardless of country. Hedstrom, Akerstedt, Hillert, Olsson and Alfredsson (2011) have related development of MS to shift work at an early age and resulting disruption to circadian rhythms. A further causal candidate is infection with the Epstein-Barr virus (Santon et al., 2011).

Epidemiological data exist that support both genetic and environmental factors in causing the condition of MS. Pugliati, Sotgiu and Rosati (2002) carried out a review of epidemiological studies from across the globe and concluded that prevalence rates

may reflect migratory patterns of populations who were already at higher genetic risk of the disease.

Cognitive Effects

While MS is popularly perceived as producing physical impairments, there has been an increasing acknowledgement in the literature that a variety of psychological difficulties are also caused by the action of the disease. There is some recent evidence of personality changes in MS sufferers associated with cognitive impairment (Benedict, Priore, Miller, Munschauer & Jacobs, 2001). Recent studies (Ouellet et al., 2010; Banati et al., 2010) have also related executive function to differences between MS patients and controls on tests of social cognition and theory of mind.

Cognitive impairment was found by Rao, Leo, Bernardin and Unverzagt (1991a) to affect around 43% of those with a diagnosis of MS. Almost a third of MS patients were impaired on measures of recent memory and a quarter were impaired on measures of sustained attention and verbal fluency. Around 20% were impaired on measures of conceptual reasoning and visuospatial perception. These impairments can be as disabling as the physical manifestations of the illness, making engagement in work and social activities less likely thus reducing quality of life (Rao et al., 1991b; Barak & Achiron, 2006). Kalmar, Gaudino, Moore, Halper and DeLuca (2008) assessed everyday functioning in Multiple Sclerosis using a standardised tool and found statistically significant differences between MS patients with and without cognitive impairments and healthy controls. Julian, Vella, Vollmer, Hadjimichael and Mohr (2008) found that around 56-58% of MS patients were not employed and this was associated with progressive course and an increase in symptoms over the previous six months. Glanz et al. (2010) found mild correlations between information processing

speed and quality of life in MS patients. Cognitive impairment has also been shown to relate to medical decision making (Basso et al., 2010) and driving ability (Schultheis, Garay & DeLuca, 2001).

Depression is also frequently found in MS sufferers with a twelve-month prevalence of 15% (Ziemsen, 2009). Studies have related depression to levels of disability (Galeazzi et al., 2005). In a longitudinal analysis over a period of seven years, Beal, Stuifbergen and Brown (2007) found that, while initially age, longer time since diagnosis and a progressive subtype of MS were associated with depression levels, only degree of functional impairment predicted depression at all periods. Ziemsen (2009) recommended ongoing screening for depression in MS. Mohr, Hart, Julian and Tasch (2007) found that asking two questions about mood and anhedonia correctly identified 99% of MS patients who had a major depressive disorder.

Factors associated with cognitive impairment

Early research assumed that cognitive impairments in MS sufferers appeared in later stages of the disease, but it is now understood that impairments can appear very early. Amato et al. (2010) recommended that neuropsychological testing should be carried out early in the disease course and regularly used to monitor levels of impairment. Testing should start as early as possible - Zipoli et al. (2010) found that in patients with a clinically isolated syndrome (a condition where MS-like symptoms are seen but MS has not yet been confirmed by Magnetic Resonance Imaging data) the presence of cognitive impairment was a predictor of later progression to full-MS diagnosis.

Julian (2011) provided an overview of the specific areas of cognition affected by MS. Reduced processing speed appears to be the most frequently observed domain

affected and can influence other areas such as working memory and executive function. However, memory impairment has also been found to be a highly prevalent deficiency (Rao, 1995), with visual and verbal episodic memory the next most frequently damaged functions. Memory impairments are found with both verbal and nonverbal information but recognition of information seems to be either normal or less impaired than other areas (Rao, 1995; DeSousa, Albert & Kalman, 2002). Difficulties are also seen in domains of attention, perceptual abilities and language. It is, however, rare for aphasia, agnosia or apraxia to be found in MS sufferers. A recent meta-analysis of studies investigating cognitive dysfunction in Relapsing-Remitting MS (Prakash, Snook, Lewis, Motl & Kramer, 2008) found that age and gender both had significant bearing on the relationship between the MS and cognitive impairment. The largest effects on cognitive function were seen in domains of motor functioning, mood, memory and learning. Cognitive impairments have often been found to be only mildly correlated with levels of physical disability. Ruggieri et al. (2003) found that cognitive decline correlated with illness duration but not disability score as measured by Kurtzke's Expanded Disability Status Scale (EDSS; Kurtzke, 1983) and studies have frequently found high levels of cognitive dysfunction even in those with low levels of disability (Haase, Lienemann & Faustmann, 2008).

Risk factors for increased cognitive impairment include psychiatric distress, pain and fatigue, all of which are frequent symptoms of MS. Barak and Achiron (2006) surveyed the frequency of fatigue in MS, finding it to occur most of the day in 30.9% of their sample and to persist for more than a year in 69% of the population. High scores for fatigue correlated significantly with low scores on a quality of life questionnaire. However, age, gender, disease duration and score on the EDSS were not correlated to levels of cognitive fatigue.

Neuroimaging and neurophsysiological studies have linked these cognitive impairments to neurological changes resulting from MS disease action. Calabrese et al. (2010) derived measures of cortical thickness from MRI images obtained from MS patients with Relapsing-Remitting course and healthy controls. Cortical thickness was lower in frontal and temporal regions for cognitively normal MS patients compared to healthy controls. In moderate and severely impaired MS patients, more widespread cortical thinning was found. Arrondo et al. (2009) linked impairments to a measure of brain synchronisation suggesting that disease action on long white matter tracts in the brain may produce cognitive impairments. Bellmann-Strobl et al. (2009) showed that changes in scores on the Paced Auditory Serial Addition Test (PASAT) – a measure of working memory and processing speed frequently used with MS patients, correlated with changes in gadolinium-enhancing lesions found on MRI scans separated by a ten-week period. Use of a rating scale for MRI images by Chamelian, Bocti, Gao, Black and Feinstein (2005) was able to identify two groups with 'high' or 'low' lesion scores. There were statistically significant differences between the scores obtained by these two groups on tests of verbal and visuospatial memory and on an overall index of global cognitive functioning.

Neuropsychological testing

Studies have found prevalence rates for cognitive impairments in MS which vary between 40-70% although there is an increasing consensus to quote rates of approximately 50%. Impairments can be highly disabling to MS sufferers and it is therefore important that they can be identified accurately and sufferers advised by qualified neuropsychologists on management or rehabilitation strategies.

Neuropsychological tests need to have qualities of validity and reliability. Validity is the ability of a test to correctly measure the cognitive skill or ability it claims to assess and is usually established by comparison to an established measure known as a 'gold standard'. Reliability is the ability of the test to produce the same measurement of the ability with the same person at different times or with different examiners. This is established by retesting the same person with the measure to see if there is a change in scores over time.

Additionally, neuropsychological tests may be used to identify people as 'impaired' or 'not impaired'. This requires prior definition of a 'cut-off score' above or below which people can be assigned to a category. When tests are used in this way there is the possibility of incorrectly assigning a cognitively preserved person as impaired (a false positive) or categorising an impaired person as cognitively preserved (a false negative). Tests can thus be described in terms of their sensitivity – the ability of a test to correctly allocate an impaired person as test positive – and specificity – the ability of a test to assign a cognitively preserved person as test negative. By plotting how sensitivity and specificity change as the 'cut-off score' is varied a Receiver Operating Characteristics (ROC) curve is obtained. The area under this curve is a measure of the accuracy of the test – how often it correctly assigns a person as either impaired or preserved.

Many tests have been used by neuropsychologists to identify MS patients with cognitive impairment. Initial research focussed on producing batteries of tests which would investigate the range of domains typically impaired by the illness. Two of these batteries are still in wide use – the Brief Repeatable Battery of Neuropsychological Tests (BRB-N) and the Minimal Assessment of Cognitive Function in Multiple

Sclerosis (MACFIMS). A summary of the tests included in each battery is provided in Table 1 below:

Table 1: Summary of tests used in two major test batteries used with Multiple Sclerosis

Brief Repeatable Battery of Neuropsychological Tests (BRB-N) – Rao et al. (1991a)	**Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) – Benedict et al. (2002)**
Paced Auditory Serial Addition Test (PASAT) Bushke Verbal Selective Reminding Test 7/24 Spatial Recall Test Controlled Oral Word Association Test (COWAT)	Paced Auditory Serial Addition Test (PASAT) Symbol Digit Modalities Test (SDMT) California Verbal Learning Test – II (CVLT-II) Brief Visuospatial Memory Test – Revised (BVMT-R) D-KEFS Sorting Test Judgement of Line Orientation Test (JLO) Controlled Oral Word Association Test (COWAT)

The BRB-N was developed for use with Multiple Sclerosis by Rao et al. (1991a) in order to establish levels of prevalence for cognitive impairments in the MS population. Rao et al. (1991a) administered a wide and comprehensive range of neuropsychological tests to 100 MS patients and 100 demographically matched controls. Of the 26 tests administered the four in the BRB-N produced the most sensitive test indices and were selected on this basis to produce a brief, repeatable battery. The BRB-N successfully distinguished cognitively impaired MS patients from matched controls with a sensitivity of 71% and a specificity of 94%. It can be administered in twenty minutes and is frequently used in studies as a gold standard measure of cognitive status. Other brief, repeatable batteries have also been developed (Beatty et al., 1995a; Basso, Beason-Hazen, Lynn & Bornstein, 1996) but these are all limited in that they only cover a narrow range of cognitive domains and therefore omit others for the purpose of brevity. As a result the MACFIMS was developed by a

consensus panel of neuropsychologists from the United States, Canada, United Kingdom and Australia (Benedict et al., 2002). It takes 90 minutes to administer and contains tests chosen for their ready availability. Each of the cognitive domains typically affected in MS is assessed by at least one test in the battery. In validation by Benedict et al. (2006), it was found to distinguish impaired MS patients from matched controls. The battery also distinguished Relapsing-Remitting patients from those with Secondary Progressive course.

However, batteries such as the BRB-N and MACFIMS cannot be routinely used in neurological clinics to identify those patients who have cognitive impairments due to the time they take to administer. They also require trained and qualified staff to use and interpret them. Shorter tests, which could be administered in less than five minutes, will be less aversive for MS patients and also could be repeatedly administered in order to monitor change in cognitive status over repeated appointments at a clinic. Researchers have therefore sought to investigate short, screening tests which can be used to identify patients at clinics who may need further, more extensive testing by a qualified neuropsychologist. The established research methodology in the literature is to compare a potential screening tool against an established gold standard such as the BRB-N or MACFIMS to derive values of sensitivity, specificity and accuracy for the screening tool.

Given that impairment is most common in the domain of processing speed and this is thought to underlie other impaired domains, the PASAT has been traditionally used as part of the Multiple Sclerosis Functional Composite for this purpose. Parmenter et al. (2007) recommended further investigation of the PASAT as a screening test, although Younes et al. (2007) found that, while it had good specificity, it demonstrated very poor sensitivity.

Given the clinical importance of early identification of cognitive impairments in MS patients or those with Clinically Isolated Syndromes, Deloire et al. (2006) investigated which neuropsychological tests best identified difficulties in a sample of newly diagnosed patients with Relapsing-Remitting disease type and concluded that the SDMT was most ideal for this purpose. When investigating the SDMT as a screening tool, Parmenter et al. (2007) found it to have sensitivity of 0.82, specificity of 0.60 and a high level of accuracy of 0.84. They recommended that future research should consider the joint use of the SDMT and another screening tool – the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ).

Development of the MSNQ is outlined by Benedict et al. (2003a). Using Rasch analysis on 80 statements generated from a survey of both the literature and professionals working with MS patients, they derived a questionnaire consisting of fifteen statements that can be self-administered in the clinic to patients or informants who know them well. Validation studies of the MSNQ have found that the informant version is a more valid measure of impairment since the patient version scores are confounded by measures of depression (Benedict et al., 2004; Vanotti et al., 2009), although O'Brien et al. (2007) and Akbar et al. (2010) also found poor values of sensitivity and specificity for both versions of the test.

Reliability of joint use of the MSNQ and SDMT as screening tests was investigated by Benedict et al. (2008). They both showed strong test-retest reliability over repeated presentations. While there were minimal practice effects in both tests, these were more evident in the SDMT than the MSNQ. These findings were also supported by Morrow et al. (2010) using a sample of 660 MS patients across 21 countries tested at monthly intervals over a 48-week period. Test-retest correlations for the SDMT between each pair of successive monthly tests ranged from 0.89 to 0.96,

while for the MSNQ, these were between 0.82 and 0.93. Morrow et al. (2010) concluded that the two tests are suitable for regular use in monitoring cognitive function in patients over time. However, they acknowledged that since no reference criterion was applied, there was no way of knowing whether the changes in scores over the 48-week period reflected real changes or whether they were due to participant practice effects and experimenters becoming more proficient at administration.

Further Research

The current study aimed to add to the literature on screening tests in MS by establishing sensitivity, specificity and accuracy values for the MSNQ and SDMT in comparison to the MACFIMS as a reference criterion.

Previous studies have assessed screening tools individually but since the current study sought to validate the joint use of two such tests, there is a possibility of assessing their combined score. McIntosh and Pepe (2002) demonstrated that an optimal way to combine two disease markers is to use logistic regression to fit a model describing the relationship of the markers to the binary dependent variable of 'diseased/not diseased'. While McIntosh and Pepe's (2002) work is mainly in the area of cancer screening, their principles apply equally well to the use of screening tests for any other binary state of illness. Using logistic regression to produce a combined score for each version of the MSNQ in association with the SDMT, values of sensitivity, specificity and accuracy can be derived for the joint use of the tests. Based upon previous research (Benedict et al., 2004; O'Brien et al., 2007) it would also be expected that MSNQ-P scores would correlate to measures of depression but that the MSNQ-I scores would not.

Hypotheses

Hypothesis 1: B-values obtained from a logistic regression on MSNQ-P and SDMT scores will be significantly related to impairment status in MS patients undergoing 'gold standard' testing.

Hypothesis 2: B-values obtained from the logistic regression on MSNQ-I and SDMT scores will be significantly related to impairment status in MS patients undergoing 'gold standard' testing.

Hypothesis 3: The area under the curve generated by ROC analysis on the combined MSNQ-P and SDMT scores will approach 1.000

Hypothesis 4: The area under the curve generated by ROC analysis on the combined MSNQ-I and SDMT scores will approach 1.000

Hypothesis 5: There will be a statistically significant positive correlation between scores on the MSNQ-P and scores on the BDI-FS.

Hypothesis 6: A statistically significant positive correlation will be found between score on the MSNQ-I and the BDI-FS.

Methodology

Design

The study was an examination of the criterion validity of the combined use of the MSNQ and SDMT to detect cognitive dysfunction in patients with Multiple Sclerosis. This was carried out using a reference criterion of the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) battery of neuropsychological tests which examines a broad spectrum of the cognitive functions usually affected by MS. The study comprised a cross-sectional single point sample correlational design.

Logistic regression was used to establish weightings for the two screening test scores in order to obtain a combined score for the two tests. Using Receiver Operating Characteristics (ROC) analysis as outlined by Pintea and Moldovan (2009), optimum values of sensitivity, specificity and accuracy were derived for each test individually and for their use in combination.

Participants

To calculate the number of participants required for adequate power the method of Peduzzi, Concato, Kemper, Holford and Feinstein (1996) was used. Peduzzi et al. (1996) established that for results of logistic regression to be meaningful the number of participants in the sample who are 'test positive' should be at least 10 times the number of independent variables under analysis. In this study there are two independent variables for each logistic regression – MSNQ score and SDMT score – so a total of 20 impaired participants were required. Given the previously reported prevalence of

impairment in MS populations of 50% this meant a sample of 40 ‘gold standard’ tested participants would be required.

Participants were recruited through Multiple Sclerosis clinics held at a local hospital and at drop in surgeries held by MS specialist nurses at a variety of community-based resource centres for people with MS. Potential participants were provided with an Information Sheet (Appendix G) and Consent Form (Appendix H) for the study, by post from the MS clinic, in advance of their attendance at their appointment with an MS nurse or consultant neurologist. Information was also provided to attendees at MS drop-in centres by MS specialist nurses who regularly attended there. The researcher then attended the drop-in centre or MS Clinic in order to obtain written consent and carry out the first screening stage of the study. A total of 112 participants were recruited and underwent testing using the two screening tools. Each recruited participant was assigned a sequential identifying number by the investigator. From this initial group, a sub-sample of 33 participants were randomly selected to undergo further ‘gold standard’ testing. Random selection was carried out by the investigator’s supervisor, who was blind to the identities of participants, using random number tables. This provided the investigator with a set of numbers that were used to refer back to individual participants who were then invited back, either by telephone or letter, to attend for second stage testing. Randomisation was set up so that 1 in 3 of the stage one participants were invited to return for stage two MACFIMS testing. This ratio was selected on the basis that it provided a balance between two factors. If the ratio was too low (ie. Recruiting every participant or 1 in 2 participants) then order effects or biases in the stage one sample could still be present in the stage two sample. If the ratio was too high then this would make the study impractical to carry out in the time available since very large numbers of stage one participants would then be required.

Of the 33 selected participants, a total of 26 attended for second stage testing with the MACFIMS battery. Participants received no financial reward for involvement.

Inclusion criteria for patient participants were:

- aged eighteen years or over
- having a confirmed diagnosis of Multiple Sclerosis made by a qualified neurologist
- all of the clinical subtypes of Multiple Sclerosis

Exclusion criteria were developed to ensure that patient participants were capable of completing all aspects of testing and that findings would not be confounded by other clinical conditions. These criteria excluded the following types of patients from the study:

- Those with a history of any other neurological disease (two potential participants were excluded for this reason)
- Those with a history of alcohol or drug dependence (none of the potential participants were excluded for this reason)
- Those with severe visual or motor impairments (one potential participant was excluded for this reason)
- Those who could not speak and understand English fluently (none of the potential participants were excluded for this reason)

When participants had a person who knew them well and could complete the informant version of the MSNQ, they were also recruited to the study and asked to complete the MSNQ-I separately from the patient participant. Inclusion criteria were used for informants although these did not need to be as strict as those for participants:

- aged over eighteen years
- to have known the patient well
- to be able to read and understand English fluently

No specific exclusion criteria were employed for informant participants.

Measures

The two screening tests used in the study were the Multiple Sclerosis Neuropsychological Questionnaire (Benedict et al., 2003a) and the Symbol Digit Modalities Test (Smith, 1982). Choice of these tests was informed by a review of the literature which found that their joint use was recommended by Parmenter, Weinstock-Guttmann, Garg, Munschauer and Benedict (2007), although no studies had formally evaluated their effectiveness when used together.

The Multiple Sclerosis Neuropsychological Questionnaire (MSNQ; Appendix K) is a short self-completed questionnaire which consists of 15 statements concerning possible cognitive difficulties which MS sufferers may experience. The test consists of two parallel versions for patients and informants who know them well. Participants rate each statement in terms of how frequent the problem concerned has been for them and how disruptive it has been over the last three months. Participants provide a score between 0 and 4 for each of the statements, with higher scores indicating higher frequency and disruptiveness. The score for the test is the total of all ratings provided across the 15 statements. The MSNQ has been shown to have good internal reliability (Benedict et al., 2003a) with a Cronbach's alpha of 0.93 for the patient version and 0.94 for the informant version. The MSNQ has been shown to distinguish MS sufferers from

healthy controls and test-retest reliability is also high (Benedict et al., 2004) with correlations of 0.90 for the patient version and 0.93 for the informant version.

The Symbol Digit Modalities Test (SDMT; Appendix L) presents participants with a set of abstract symbols which are paired with a single digit between one and nine. These are given in a key at the top of the stimulus sheet and participants are required to provide the corresponding digit for a sequence of the abstract symbols provided. The SDMT was administered orally as recommended by previous studies (Rao, Leo, Bernardin & Unverzagt, 1991a; Parmenter et al., 2007) where participants provide the digit response verbally rather than in written form. The researcher recorded the sequence of numbers provided and the score for the test was the number of correct responses provided in a 90-second period. Smith (1991) reports good test-retest reliability correlations (r=0.76) for the oral version of the test. Validity is demonstrated by high correlation with the digit symbol substitution test (r=0.91; Morgan & Wheelock, 1992) and factor analysis studies showing the SDMT measures the same aspects of attention as the Test of Everyday Attention (Bate, Matthias & Crawford, 2001).

The choice of tests for the neuropsychological battery used in the second stage of the study was informed by the established literature and followed that recommended in the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS; Benedict et al., 2002). This is a battery of tests devised by a consensus panel of experts and incorporates tests which cover the range of impairments usually associated with MS. In addition, Engel, Greim and Zettl (2007) recommend that tests should take account of the impact of both fatigue and depression on cognitive abilities. In the case of depression, an adapted version of the established measure is available which omits items which would be confounded by symptoms typical in MS. The battery used is

designed to take no longer than 90 minutes to complete in full with a participant (Benedict et al., 2002). The MACFIMS consists of the following tests:

1. Judgement of Line Orientation (Benton, Hamsher, Varney & Spreen, 1983)
2. Controlled Oral Word Association Test (Benton & Hamsher, 1989)
3. California Verbal Learning Test – II (Delis, Kramer, Kaplan & Ober, 2000)
4. Brief Visuospatial Memory Test – Revised (Benedict, 1997)
5. Delis-Kaplan Executive Functioning System – Sorting Test (Delis, Kaplan & Kramer, 2001a)
6. Beck Depression Inventory – Fast Screen (Beck, Steer & Brown, 2000)
7. Paced Auditory Serial Addition Test (Gronwall, 1977)
8. Fatigue Impact Scale (Fisk et al., 1994)

The MACFIMS battery has been validated for use in MS by Benedict et al. (2006). All the tests contained in the battery effectively discriminated MS patients from normal controls with effect sizes ranging from 0.49 to 1.31 for the SDMT ($p<0.005$). The MACFIMS also distinguished MS patients with relapsing-remitting course from those with a secondary-progressive disease.

With the exception of the MSNQ and the FIS, all of the tests used in the current study were already used as part of routine practice in the neuropsychology department where testing was carried out. Permission to use the MSNQ was obtained over

electronic correspondence with its author Dr. Ralph Benedict, Ph.D. Permission to use the FIS was obtained from MAPI Research Trust.

Psychometric Tests Used in MACFIMS

The Judgement of Line Orientation Test (JLO, Appendix M; Benton et al., 1983) is a measure of visuospatial perception. The participant is presented with a stimulus consisting of two lines at specific orientations in the top half of the display booklet and a multiple choice array of eleven lines labelled with the numbers one through to eleven (which remains the same for all stimuli) in the bottom half of the display booklet. Participants are asked to tell the tester the numbers corresponding to the two lines in the multiple choice array that are in the same position and point in the same direction as the two stimulus lines. There are five practice items followed by thirty test items. The test has two equivalent forms – form H and form V which both contain the same items displayed in a different order with generally ascending levels of difficulty. All participants in the current study completed form V. Internal consistency of the test is established (r=0.84 to 0.91; Benton, Sivan, Hamsher, Varney & Spreen, 1994; Qualls, Bliwise & Stringer, 2000; Vanderploeg, LaLone, Greblo & Schinker, 1997; Winegarden, Yates, Moses, Benton & Faustman, 1998; Woodward et al., 1996). Test scores correlate highly with visual spatial subtests of the WAIS-R (Block design r=0.68; Object Assembly r=0.69; Trahan, 1998).

When undertaking the Controlled Oral Word Association Test (COWAT, Appendix N; Benton & Hamsher, 1989) participants are asked to provide as many words as they can which begin with each of three letters of the alphabet within a one minute period. The test was chosen for the MACFIMS for its excellent psychometric

properties and sensitivity to assessing the language impairments seen with MS. The test is administered using various sets of three letters but versions using F/A/S have been shown to have high correlation coefficients with the standard version using C/F/L or P/R/W (r=0.97-0.94; Lacy et al., 1996). The test also has extremely high inter-rater reliability (r=0.98; Norris, Blankenship-Reuter, Snow-Turek & Finch, 1995) and strong test-retest reliability (r=0.74; Ruff, Light, Parker & Levin, 1996).

The California Verbal Learning Test –II (CVLT-II, Appendix O; Delis et al., 2000) is a psychometric measure of verbal memory skills. Participants are presented aurally with a standardised list of 16 words and asked to recall as many as possible. Participants are allowed to provide free recall of the words. Presentation is repeated after each participant recall until there has been a total of five presentations of the word list. Participants are then given a distraction task consisting of the aural presentation of a new sixteen item word list which they are asked to recall before being asked for free recall of the original list. Since each of the word lists consists of sets of four words from four different conceptual categories, a further cued recall task is also then given for the original word list. After a delay of at least 20 minutes, further free and cued recall trials are carried out. A recognition task is also administered requiring participants to recognise the 16 words from the original list from an aurally presented list of 48 words, which also contains the sixteen words from the distraction list and new words which are both conceptually linked and distinct from the original list. The CVLT-II has an alternative equivalent form but this was not used for the current study. Delis et al. (2000) reported test-retest correlations above 0.80 for the total correct scores over all five learning trials and the long-delay free recall score. Validation of the test comparing it to the first version of the CVLT found a correlation coefficient of 0.76 for the total recall across the learning trials.

To assess visuospatial memory and learning, the MACFIMS panel recommended use of the Brief Visuospatial Memory Test – Revised (BVMT–R, Appendix P; Benedict, 1997). This test has six equivalent forms of which Form 2 was used here. Participants are presented with a 2 x 3 matrix of symbols on which learning is assessed over three presentations lasting 10 seconds each, with a delayed recall trial and recognition trial after a 25-minute period has elapsed from the learning trials. If participants have problems in manual dexterity there is an optional copy trial (where the stimulus is displayed while the participant copies it), which can be administered after completion of the recognition trial in order to assist in marking responses. Benedict (1997) found the test-retest reliability coefficient for the total recall score on the BVMT-R to be 0.80 and inter-rater reliability for marking the test was also high – being greater than 0.90. Studies of validity (Benedict, Schretlen, Groninger, Dobraski & Shpritz, 1996) have shown that BVMT-R scores for learning and delayed recall correlate strongly with other tests of explicit memory such as the visual reproduction subtest of the Wechsler Memory Scale - Revised and the Rey Complex Figure test ($r=0.65 – 0.80$).

The MACFIMS includes the Delis-Kaplan Executive Function System - Sorting Test (DKEFS-ST, Appendix Q; Delis, Kaplan & Kramer, 2001a) in order to assess executive function and it specifically tests a participant's conceptual reasoning skills. The examinee is presented with a set of six cards which can be sorted in a variety of ways based on both conceptual and perceptual criteria. They are asked to sort these cards into two groups that each contain three cards and explain the concepts used to generate their sorting decision; they repeat this until they cannot generate any further groupings. For each form there are two sets of cards for which this is done. There are two equivalent forms for the test but the current study used card sets one and two. The

test also contains a sort recognition condition but the MACFIMS panel do not recommend administration of this condition in order to reduce administration time. This advice has been followed for the current study. Internal consistency values for the first condition of the test in adults are established (r=0.72-0.86; Delis et al., 2001b). Correlations between all DKEFS scores and the Wisconsin Card Sorting test have been found to be moderate to high (r=0.31 to 0.59; Delis et al., 2001b).

Completion of the Paced Auditory Serial Addition Test (PASAT, Appendix R; Gronwall, 1977) requires that participants focus their attention and are also able to carry out rapid information processing. As such, it is a test of both processing speed and working memory. Examinees are required to monitor a series of 61 single digit numbers provided from audiotape or CD, add the number just presented to the one immediately preceding it and verbally report the total. Digits in the original version (Gronwall, 1977) were presented at varying inter-stimulus intervals. However, a version modified for use with MS patients was later developed by Rao et al. (1991a) where the stimulus intervals were 3.0 seconds and 2.0 seconds. It is this version of the PASAT that was used for the current study. The test has high internal reliability across versions (r=0.90; Crawford, Obansawin & Allan, 1998) and strong test-retest reliability (r>0.90; McCaffrey et al., 1995). Construct validity is demonstrated by correlations to scores on other tests of attention such as the digit span and arithmetic subtests of the WAIS-R (Tombaugh, 2006). The test is widely used with MS patients as part of the Multiple Sclerosis Functional Composite (Fischer, Jak, Kniker, Rudick & Cutter, 2001).

In line with the recommendations of Engel et al. (2007), two questionnaires measuring depression and fatigue were also included in the present battery. The tool used for measuring depression was the Beck Depression Inventory – Fast Screen (BDI-FS, Appendix S; Beck et al., 2000). Participants are presented with 7 items on the

questionnaire consisting of 4 statements related to depression. They are asked to select which of these four best describes how they have been feeling over the last 2 weeks. Each of the statements selected is scored between 0 and 3 to give a total score for the participant out of a total of 21. The BDI-FS has been validated as a measure of depression in an MS sample by Benedict, Fishman, McClellan, Bakshi and Weinstock-Guttman (2003b), finding it to be significantly correlated with other measures of depression such as the Centre for Epidemiological Studies Depression questionnaire (CES-D, r=0.86).

The Fatigue Impact Scale (FIS, Appendix T; Fisk et al., 1994) consists of a set of 40 statements describing the effects of fatigue that people experience. The participant is asked to rate each of them on a Likert scale in order to describe how much of a problem fatigue has been for them over the past 4 weeks. The scale provides scores out of 40 for cognitive and physical fatigue as well as a score out of 80 for social fatigue. The scale has been validated in an MS population by Mathiowetz (2003) who demonstrated it to have moderate correlations with subscales of the SF-36 health survey (r=-0.56 – 0.62) and good test-retest reliability (Intraclass correlation coefficient = 0.68 – 0.85).

Procedure

The study was carried out in two stages. Stage one involved the completion of the patient-version MSNQ and the SDMT with patients and the informant-version MSNQ by their relative/carer during their attendance at routine MS clinics at a large general hospital. For the purposes of gaining consent for the study Information Sheets and Consent Forms were sent out from the MS Clinic in advance of their appointments

to both the clients and their informants. The researcher attended regular MS Clinics to explain the study to potential participants in more detail, obtain their informed consent and carry out the screening tests. A sample of 112 participants was recruited to this stage of the study.

A sub-sample of 33 participants for stage two of the study was selected using random number tables giving every participant from stage one a 1 in 3 probability of being selected. This second stage involved the completion of the MACFIMS standardised battery of neuropsychological tests with participants by the experimenter at the clinical neuropsychology service at Leicester General Hospital. The test battery was administered in a set order which allowed for the required delay between initial learning and delayed recall trials for the BVMT-R and CVLT-II. The order of administration is outlined below:

1. Brief Visuospatial Memory Test – Revised: Learning Trials 1-3
2. California Verbal Learning Test – II: List A Trials 1-5, List B, Short Delay List A Free Recall and Cued Recall
3. Delis-Kaplan Executive Function System – Sorting Test: Free Sorting Card Sets 1 & 2
4. Brief Visuospatial Memory Test – Revised: Delayed Recall and Recognition Trials, Copy trial if required
5. California Verbal Learning Test – II: Delayed Recall List A Free Recall and Cued Recall, Recognition Trial
6. Judgement of Line Orientation Test
7. Beck Depression Inventory – Fast Screen
8. Fatigue Impact Scale
9. Controlled Oral Word Association Test – F/A/S

10. Paced Auditory Serial Addition Test: Form A

Testing was carried out in a quiet, private and comfortable room allocated by the service as suitable for neuropsychological testing. It was always ensured that a ground floor room with disabled access was made available to the researcher so that mobility problems did not prevent involvement of any participants.

Ethical Considerations

The study was designed and implemented in line with the British Psychological Society Code of Human Research Ethics (British Psychological Society, 2010). Ethical approval for the study was obtained from a local research ethics committee and from the research and development department of the hospital where the data collection occurred. Letters concerning this ethical approval process can be found in Appendix U.

Informed consent was obtained from all participants using standardised Information Sheets and Consent Forms (Appendices G - J) explaining the aims of the study, what the data collected would be used for and informing participants that they could withdraw consent to take part at any point. The principal researcher was available by telephone to discuss any aspects of the project participants wished to enquire about. Consent Forms and Information Sheets were constructed in line with guidance from the National Research Ethics Service.

The Paced Auditory Serial Addition Test (PASAT) is known to produce some distress in participants if it highlights cognitive difficulties. Given that the current study aimed to test for such difficulties, it was considered likely that some participants would experience distress during this part of the neuropsychological battery. In order to reduce

distress, the PASAT was introduced as “sometimes being a difficult test for people” in order to normalise this experience. It was also administered last in the neuropsychological battery so that time to allay any distress was available immediately after completion.

Results

A total of 112 participants were recruited from Multiple Sclerosis clinics at a local hospital and drop in surgeries based at community-based resource centres. From this group, a total of 33 participants were independently randomly selected to take part in the second stage of the study to complete the MACFIMS battery of tests. However, of these, a total of six participants were unable to attend their appointment within three months of completing the screening either due to not responding to contacts made inviting them to a further appointment or due to illnesses that prevented them being able to come to the testing facility. After being invited to stage two of the research, a further participant contacted the experimenter to withdraw from the study and their data were removed from the analysis.

Therefore the final sample of data analysed constituted a total of 26 patient participants who had completed both stage one and stage two of the study. The average age of participants in the sample was 49.5 years (S.D.=10.9 years; range = 31.1 – 81.1 years). Mean education was 13.7 years (S.D. = 3.4 years; range = 10.0 – 20.0 years). Mean number of years since receiving a diagnosis was 8.6 years (S.D. = 8.3 years; range = 0.5 – 42.0 years). Fifteen participants were female (57.7%) and eleven were male (42.3%). Twenty one participants (80.8%) described themselves to be of White-British ethnicity, three (11.5%) described themselves as Asian British, one participant (3.8%) identified as White-Other and one (3.8%) identified as of Mixed ethnicity. The sample consisted of fifteen (57.7%) participants with a Relapsing-Remitting diagnosis, five (19.2%) with a Secondary-Progressive diagnosis and six (23.1%) with a Primary-Progressive subtype. A total of 20 patient participants provided someone who could be recruited as an informant and all of these completed the informant-version of the MSNQ.

All twenty six patient participants completed the MACFIMS battery in order to ascertain their genuine impairment status. The tests in the MACFIMS produce a wide range of scaled scores, T scores, z scores and percentile scores. In validating the MACFIMS battery, Benedict et al. (2006) identified a total of 11 of these scores, which had been shown in the literature to reliably identify impaired from non-impaired MS patients with statistically significant effect sizes. Since the SDMT is used as a screening test in the current study the total correct score for this test was not used leaving 10 headline test scores used to assign participants to the Impaired or Non-Impaired category. These headline scores were: the COWAT total score (COWAT-Tot); DKEFS-Sorting Test Correct Sorts (DKEFS-CS) and Description Score (DKEFS-DS); Judgement of Line Orientation Total Correct (JLO-Tot); Paced Auditory Serial Addition Test 3-second inter-stimulus interval score (PASAT-3) and 2-second inter-stimulus interval score (PASAT-2); California Verbal Learning Test-II Total Learning Trials Score (CVLT-TL) and Delayed Recall Score (CVLT-DR); Brief Visuospatial Memory Test-Revised Total Learning Trials Score (BVMT-R-TL) and Delayed Recall (BVMT-R-DR). Tables outlining effect sizes from previous research and from Benedict et al. (2006) are found below. The criterion for identifying any participant as Impaired was a score in the borderline range (more than 1.5 standard deviations below the norm) on any of these 10 headline test scores.

Table 2:Effect sizes and significance levels from previous studies for headline test scores in the MACFIMS battery (adapted from Benedict et al.,2006). CCST-CS and CCST-DS represent the Correct Sorts and Description Scores from the California Card Sorting Test – a precursor test to the DKEFS Sorting test.

Test Score	Study	Effect Size	Significance level
COWAT-Tot	Beatty & Monson (1994)	0.8	0.01
	Rao et al. (1991a)	0.8	<0.001
	Ryan et al. (1996)	0.4	<0.01
	Sperling et al. (2001)	1.0	<0.01
CCST-CS	Beatty & Monson (1996)	1.0	0.001
	Beatty et al. (1995b)	0.9	0.001
CCST-DS	Beatty & Monson (1996)	0.8	0.01
	Beatty et al. (1995b)	0.6	0.01
JLO-Tot	Benedict et al. (2001)	1.0	0.01
	Rao et al. (1991a)	0.5	<0.01
PASAT-3	Sperling et al. (2001)	1.1	<0.01
	Rao et al. (1991a)	0.7	<0.01
PASAT-2	Rao et al. (1991a)	0.6	<0.01
CVLT-TL	Benedict et al. (2001)	1.0	0.01
	Thornton et al. (2002)	0.7	<0.001
CVLT-DR	Thornton et al. (2002)	0.6	<0.01
BVMT-R-TL	Benedict et al. (2001)	1.1	0.01

Table 3: Effect sizes and significance levels for headline test scores in the MACFIMS test battery from the Benedict et al. (2006) validation study.

Test Score	Effect Size	Significance level
COWAT	0.53	<0.001
JLO-Tot	0.49	0.004
CVLT-TL	0.70	<0.001
CVLT-DR	0.79	<0.001
BVMT-R-TL	1.04	<0.001
BVMT-R-DR	1.07	<0.001
PASAT-3	0.61	<0.001
PASAT-2	0.58	<0.001
DKEFS-CS	0.64	<0.001
DKEFS-DS	0.69	<0.001

Using this criterion for impairment, a total of 11 participants were classified as Impaired. The prevalence level of impairment in the sample was therefore consistent with previous research at 42.3%. Of these 11 impaired participants, 7 were female (63.6%) and 4 were male (36.4%). A Chi-Squared analysis found no significant differences between the impaired and non-impaired groups with regard to gender, χ^2 (df=1, n=26) = 0.276, p=0.599.

There were 9 impaired participants with a Relapsing-Remitting course (81.8%), none with Secondary Progressive type and 2 with Primary Progressive MS (18.2%). A chi-squared test showed that there was no statistically significant difference between Impaired and Non-Impaired groups, although the p-value for this result approached significance, χ^2 (df=2, n=26) = 5.788, p=0.055.

The Impaired group was made up of 9 White-British (81.8%), 1 Asian-British (9.1%) and one mixed ethnicity (9.1%) participants. There was no significant difference between the Impaired and Non-Impaired groups in terms of ethnicity, χ^2 (df=3, n=26) = 2.199, p=0.532.

The Impaired group of participants had a mean age of 48.47 years, mean number of years of education of 12.27 and mean duration since diagnosis of 11.86 years. Comparison of the Impaired and Non-Impaired groups on these variables was carried out using an independent samples t-test. No significant differences between the groups were found for age (t (df=24) = 0.406, p=0.689).

Differences for number of years of education were not significant but the p-value for this result approached significance (t (df=24) = 2.013, p=0.055). A similar result was seen for number of years since diagnosis with a slightly higher p-value (t (df=24) = -1.803, p=0.084).

SPSS output for these results can be found in Appendix X.

Individual test ROC analysis

Receiver Operating Characteristics (ROC) analysis, as described by Pintea and Moldovan (2009), was carried out on the scores obtained on each of the tests individually. Scores on the patient version of the MSNQ produced a ROC curve with an area under the curve significantly different from chance, AUC = 0.785 (n=26, p= 0.015, std.error = 0.098, 95% confidence interval = 0.594 – 0.976). At the optimum cut-off score of 30 or above the MSNQ-P had a sensitivity of 0.727 and specificity of 0.867. The ROC curve for this test is shown below:

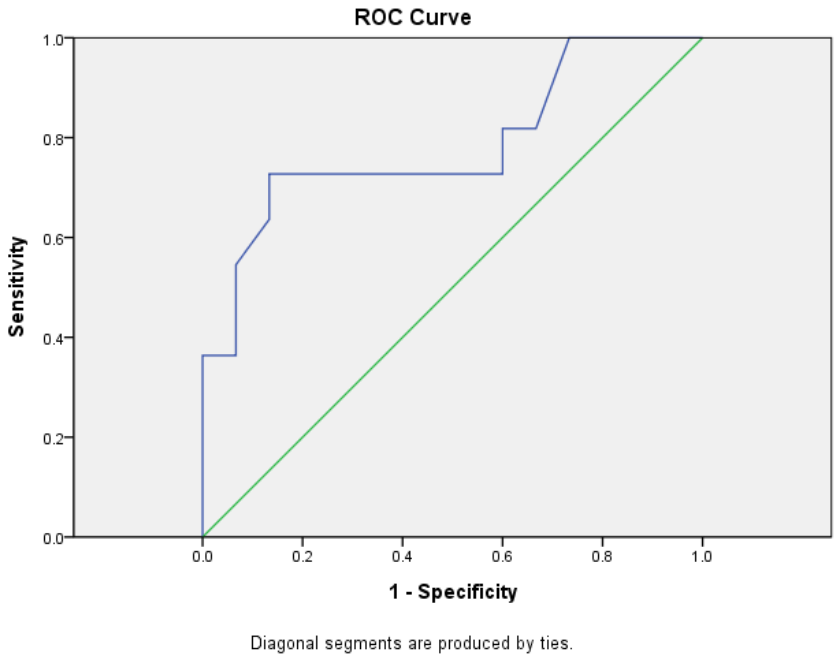

Fig.1: ROC curve for MS Neuropsychological Questionnaire – Patient Version

ROC analysis on scores obtained on the MSNQ-I produced a curve with AUC = 0.652 (n=20, p= 0.254, std.error = 0.126, 95% confidence interval = 0.405=0.898). At the optimum cut-off score of 15 or above, the MSNQ-I had a sensitivity of 0.778 and specificity of 0.455. The ROC curve for this test is shown below:

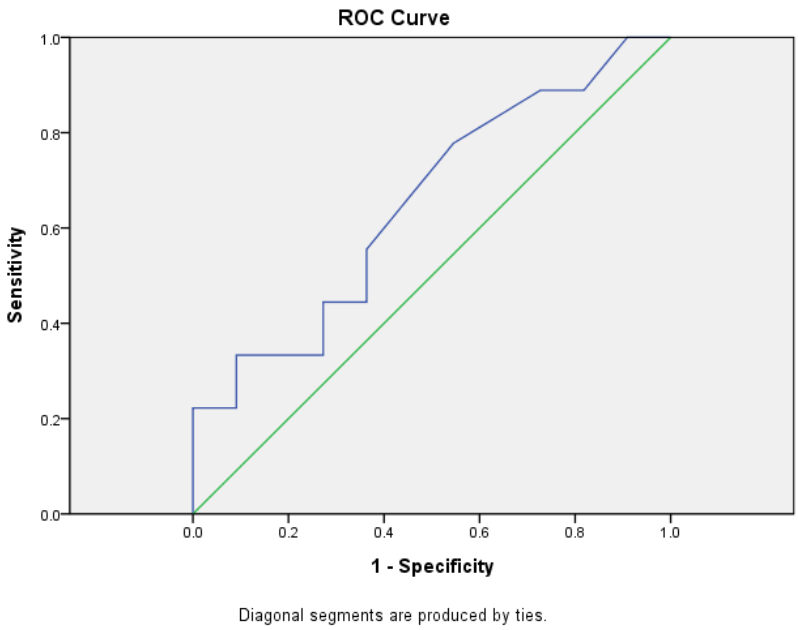

Fig. 2: ROC curve for MS Neuropsychological Questionnaire - Informant Version

ROC analysis of scores on the Symbol Digit Modalities Test (SDMT) produced the curve as shown in Figure 3 below. Because low scores on the SDMT indicate impairment, scores were inverted for this analysis by subtracting the obtained score from the maximum possible SDMT score of 110. The ROC curve is thus inverted to make it visually comparable to the curves provided from the MSNQ. The ROC curve had an AUC = 0.712 (n=26, std.error = 0.712, p = 0.069, 95% confidence interval = 0.510 – 0.914). An inverted score of 63 or above produced optimum sensitivity of 0.727 and specificity of 0.667 for the test. Therefore an optimum cut-off for the SDMT raw score is any score below 47 symbols correct in 90 seconds.

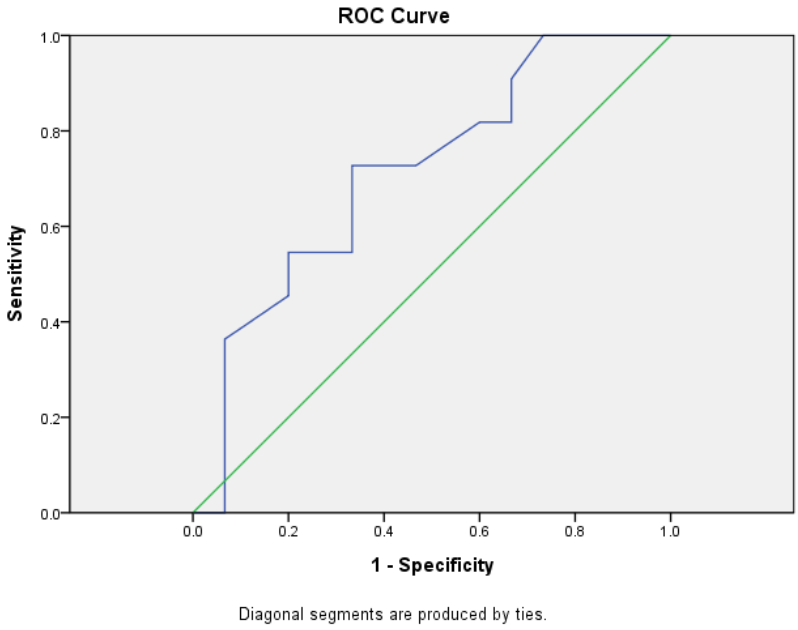

Fig. 3: ROC curve for Symbol Digit Modalities Test

SPSS output for these results can be found in Appendix Y.

Logistic Regression and Derivation of Combined Scores

Using the method of McIntosh and Pepe (2002), logistic regression was carried out on the scores obtained from each version of the MSNQ in combination with scores from the SDMT in relation to classification of each participant as 'impaired/not impaired' on the MACFIMS gold standard testing. The regression can be carried out either including or not including the interaction between the variables but it was considered by the researcher better to include interactions, since the cognitive domains assessed by the SDMT are also part of the overall assessment made by the MSNQ. The model generated from the regression can then be used to calculate a risk score indicative of the probability of impairment for each participant upon which a ROC analysis can be carried out. This risk score is effectively a combined score for the two screening tests.

Logistic regression was carried out on MSNQ-P scores and SDMT scores using the forced entry method. Models were generated to both include and exclude the interaction between the variables and results are outlined in Tables 5 and 6 below:

Table 4: Values of constants, Wald statistics, significance values and values of Exp(B) for logistic regression model fitted on scores from MSNQ-P and SDMT excluding their interaction

Variable	B-value (Standard Error)	Wald Statistic	Degrees of Freedom	Significance Value	Exp (B)
MSNQ-P	0.102 (0.051)	3.961	1	0.047	1.107
SDMT	-0.041 (0.045)	0.824	1	0.364	0.960
Constant	-1.137 (2.691)	0.179	1	0.673	0.321

Table 5: Values of constants, Wald statistics, significance values and values of Exp(B) for logistic regression model fitted on scores from MSNQ-P and SDMT including their interaction

Variable	B-value (Standard Error)	Wald Statistic	Degrees of Freedom	Significance Value	Exp (B)
MSNQ-P	0.448 (0.342)	1.720	1	0.190	1.565
SDMT	0.144 (0.170)	0.718	1	0.397	1.155
MSNQ-P by SDMT	-0.008 (0.007)	1.152	1	0.283	0.992
Constant	-9.453 (8.163)	1.341	1	0.247	0.000

The model produced, which included the interaction between MSNQ-P and SDMT scores, did not produce b-values that differed significantly from zero. The model fitted when excluding the interaction produced a b-value for the MSNQ-P score, which did differ significantly from zero, although the b-value for both the constant and SDMT score were not significantly different from zero. A pragmatic decision was made to use the model fitted without the interaction to generate the risk score for subsequent ROC

analysis. The equation used to generate a combined score for ROC analysis was therefore:

Combined Score = (0.102*MSNQ-P) + (-0.041*SDMT) – 1.137

Results of logistic regression for scores on the MSNQ-I and SDMT using the forced entry method are outlined below. As for the analysis above, models were generated to both include and exclude the interaction between the variables.

Table 6: Values of constants, Wald statistics, significance values and values of Exp(B) for logistic regression model fitted on scores from MSNQ-I and SDMT excluding their interaction

Variable	B-value (Standard Error)	Wald Statistic	Degrees of Freedom	Significance Value	Exp (B)
MSNQ-I	0.027 (0.045)	0.362	1	0.547	1.028
SDMT	-0.086 (0.061)	1.986	1	0.159	0.917
Constant	3.202 (3.291)	0.947	1	0.331	24.592

Table 7: Values of constants, Wald statistics, significance values and values of Exp(B) for logistic regression model fitted on scores from MSNQ-I and SDMT including their interaction

Variable	B-value (Standard Error)	Wald Statistic	Degrees of Freedom	Significance Value	Exp (B)
MSNQ-I	0.828 (0.590)	1.969	1	0.161	2.288
SDMT	0.234 (0.216)	1.168	1	0.280	1.263
MSNQ-I by SDMT	-0.018 (0.013)	2.018	1	0.155	0.982
Constant	-11.408 (10.391)	1.205	1	0272	0.000

Neither model produced b-values that were significantly different from zero, so a pragmatic decision was made to continue the analysis using the model that produced the best significance values. Better levels of significance for constants were obtained by

including the interaction of the test scores, so the equation used to calculate the combined score was:

Combined Score = (0.828*MSNQ-I) + (0.234*SDMT) + (-0.018*MSNQ-I*SDMT) – 11.408

SPSS output for these logistic regressions can be found in Appendix Z. Since none of the models produced a complete set of constants that were found to be significantly related to impairment status, neither null hypothesis one nor null hypothesis two can be rejected.

ROC analysis on combined scores

ROC analysis on the combined scores obtained from participant scores on both the MSNQ-P and SDMT tests produced the ROC curve as shown in Figure 4 below. The area under the curve was significantly different from chance, AUC = 0.773 (n=26, p=0.020, std. error = 0.097, 95% confidence interval = 0.583-0.963). At the optimum cut-off score of -0.1275, the combined score had a sensitivity of 0.727 and a specificity of 0.800. These values are superior to those produced by the ROC analysis on the SDMT scores individually but inferior to those obtained by individual analysis of the MSNQ-P scores.

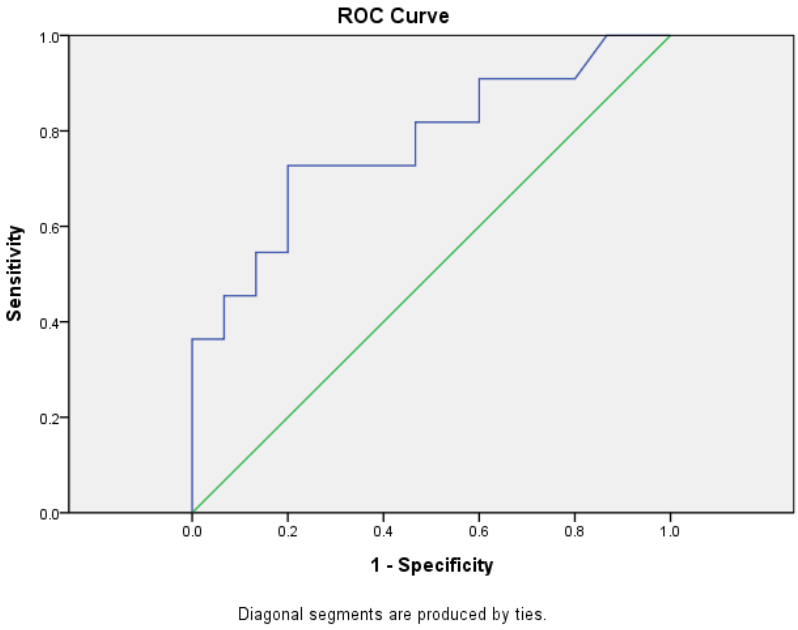

Fig. 4: ROC curve for combined score of MSNQ-P and SDMT

ROC analysis on combined scores from the MSNQ-I and SDMT tests produced the ROC curve as shown in Figure 5 below, with a superior area under the curve to that found for both tests individually, AUC = 0.808 (n=20, std. error = 0.103, p = 0.020, 95% confidence interval = 0.605 -1.000). At the optimum cut-off score of -0.7520, the combined score had a sensitivity of 0.889 and specificity of 0.636. The combined score has superior accuracy, sensitivity and specificity compared to the MSNQ-I individual score. In comparison to the SDMT individual score, the combined value had superior accuracy and sensitivity with slightly lower but roughly comparable specificity.

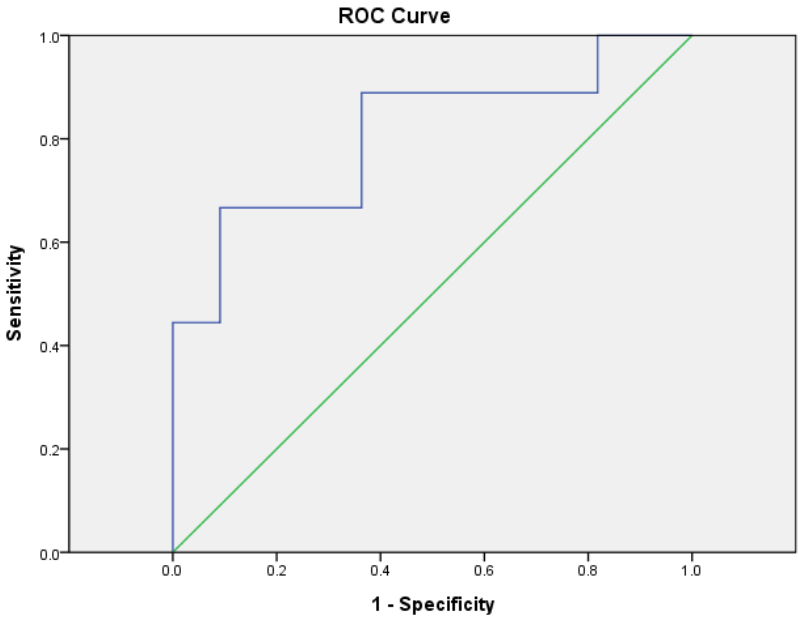

Fig. 5: ROC curve for combined score of MSNQ-I and SDMT

SPSS output for these ROC analyses can be found in Appendix AA. Both of the combined scores produced a ROC curve with an area under the curve statistically significant from that which would be expected by chance. Additionally, the process of optimising the scores by combining them has produced superior numerical values for the area under the curve compared to those obtained for the individual tests so, both hypothesis III and hypothesis IV are supported. The combination of the two tests gives a more accurate result than their use alone meaning that combined use of the MSNQ and SDMT to screen for cognitive impairments in MS clinics is recommended.

A summary of the cut-off scores, sensitivity, specificity and accuracy for each of the individual and combined tests is outlined in Table 8 below:

Table 8: Values of optimum cut-off score, sensitivity, specificity and accuracy for both versions of MSNQ and SDMT

Test	Optimum Cut-off	Sensitivity	Specificity	Accuracy
MSNQ-P	≥30	0.727	0.867	0.785
MSNQ-I	≥15	0.778	0.455	0.652
SDMT	<47	0.727	0.667	0.712
MSNQ-P & SDMT	≥-0.1275	0.727	0.800	0.773
MSNQ-I & SDMT	≥-0.7520	0.889	0.636	0.808

Analysis – Low Mood and MSNQ

The Beck Depression Inventory – Fast Scale (BDI-FS) – and Fatigue Impact Scale (FIS) were included with the gold standard battery at stage two testing since both fatigue and depression have also been related to cognitive impairment in MS by previous research. Previous research has found that MSNQ-P scores correlate significantly to measures of mood but that this is not the case for the MSNQ-I (Benedict et al., 2004; O'Brien et al., 2007).

Correlational analyses were carried out on MSNQ scores and BDI-FS scores. Since scores on these questionnaires constitute non-parametric data, the appropriate statistical test was to calculate the Spearman's rho rank order coefficient, ρ. MSNQ-P score was significantly positively correlated to BDI-FS scores, $\rho(26) = 0.684$, $p<0.01$. While the correlation coefficient for the MSNQ-I score and the BDI-FS score was not as high as for the patient version, it was still statistically significant, $\rho(20) = 0.616$, $p<0.01$. These results allow null hypotheses 5 and 6 to be rejected and the experimental hypotheses to be accepted.

Post-Hoc Analyses

Correlational analysis was also carried out for MSNQ scores with scores from the FIS. The FIS produces three domain scores – for cognitive, social and physical fatigue levels. Values of the correlation coefficient for each of these in relation to both of the MSNQ questionnaires are provided in Table 9 below:

Table 9: Values of Spearman's Rank Order Correlation Coefficient for sub-domains of the Fatigue Impact Scale with the Patient and Informant versions of the Multiple Sclerosis Neuropsychological Questionnaire

***Correlation is significant at the p<0.01 level (1-tailed test)*

	FIS-Cognitive	**FIS-Physical**	**FIS-Social**
MSNQ-P	0.858**	0.774**	0.758**
MSNQ-I	0.642**	0.573**	0.656**

Scores on both versions of the MSNQ are highly significantly correlated to scores on all of the sub-domain scores of the FIS. The correlations obtained are higher for the MSNQ-P than for the MSNQ-I and especially high between the MSNQ-P and FIS-Cognitive score. SPSS output for these analyses can be found in Appendix AB.

Given these significant correlations between both versions of the MSNQ and low mood and fatigue, it is reasonable to consider whether the BDI-FS or FIS-Cognitive domain scores could be used as screening tools and what the measures of accuracy, sensitivity and specificity would be for these. Therefore an exploratory ROC analysis was carried out on both scores.

ROC analysis using the BDI-FS total score produced a curve that did not differ significantly from what would be expected by chance: the area under the curve, AUC = 0.664 (n=26, std. error = 0.111, p = 0.161, 95% confidence interval = 0.445 – 0.882). The BDI-FS was not shown to reliably distinguish Impaired from Non-Impaired

participants. At the optimum cut-off point of scores greater than or equal to 4.5, it had a sensitivity of 0.636 and specificity of 0.667. The ROC curve produced is shown below in Figure 6:

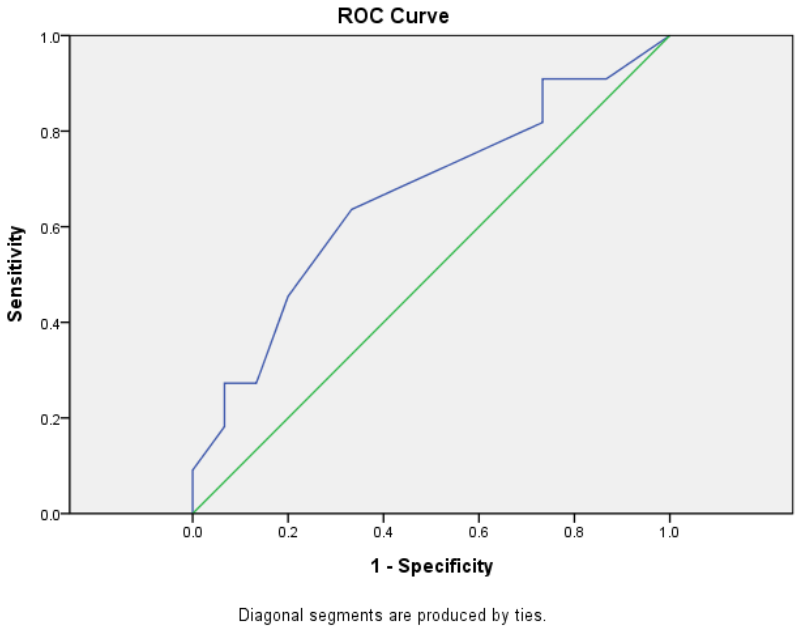

Fig. 6: ROC curve for total score on Beck Depression Inventory – Fast Screen

ROC analysis on the FIS-Cognitive score produced a curve that differed in a highly statistically significant way from what would be expected by chance. The area under the curve was higher than for any of the individual tests considered above and also higher than that found for their combined scores, AUC = 0.818 (n=26, std. error = 0.092, p = 0.006, 95% confidence interval = 0.637 – 0.999). At an optimum cut-off score of 17.5 or above, the FIS-Cognitive score had a sensitivity of 0.818 and specificity of 0.733. The ROC curve obtained is shown in Figure 7 below:

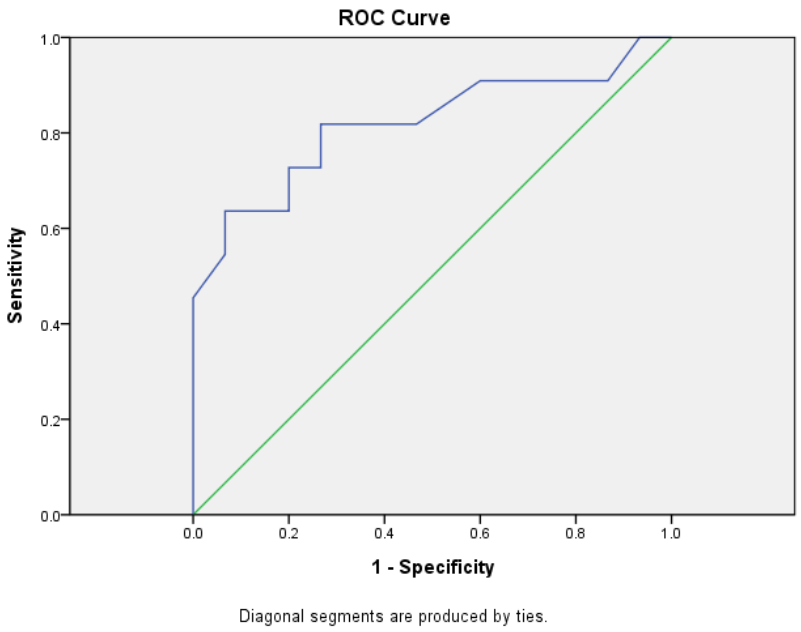

Fig. 7: ROC curve for score on Fatigue Impact Scale – Cognitive sub-domain

SPSS output for these additional ROC analyses can be found in Appendix AC.

To determine the contribution made by low mood and fatigue to the categorisation of a participant as either Impaired or Not Impaired, a further logistic regression was carried out as before, but with scores on the BDI-FS and FIS also included in the analysis. Results of these analyses are provided in Tables 10 and 11 below:

Table 10: Values of constants, Wald statistics, significance values and values of Exp(B) for logistic regression using scores from MSNQ-P, SDMT, BDI-FS and FIS

Variable	B-value (Standard Error)	Wald Statistic	Degrees of Freedom	Significance Value	Exp (B)
MSNQ-P	0.087 (0.086)	1.029	1	0.310	1.091
SDMT	-0.041 (0.071)	0.339	1	0.560	0.960
BDI-FS	0.024 (0.199)	0.015	1	0.904	1.024
FIS-Cognitive	0.231 (0.137)	2.845	1	0.092	1.260
FIS-Physical	-0.024 (0.125)	0.036	1	0.850	0.977
FIS-Social	-0.089 (0.094)	0.911	1	0.340	0.914
Constant	-1.318 (3.691)	0.128	1	0.721	0.268

Table 11: Values of constants, Wald statistics, significance values and values of Exp(B) for logistic regression using scores from MSNQ-I, SDMT, BDI-FS and FIS

Variable	B-value (Standard Error)	Wald Statistic	Degrees of Freedom	Significance Value	Exp (B)
MSNQ-I	-0.025 (0.066)	0.146	1	0.702	0.975
SDMT	-0.090 (0.101)	0.792	1	0.373	0.914
BDI-FS	0.139 (0.252)	0.305	1	0.581	1.149
FIS-Cognitive	0.152 (0.140)	1.167	1	0.280	1.164
FIS-Physical	0.158 (0.179)	0.782	1	0.376	1.171
FIS-Social	-0.107 (0.105)	1.034	1	0.309	0.898
Constant	1.164 (5.570)	0.044	1	0.835	3.202

None of the constants were statistically significant in either analysis and the standard errors associated with each constant were all as large, if not larger than the B-values. Comparison of the values of Exp (B) showed that the BDI-FS and FIS make equivalent (and sometimes greater) contributions to the model to the MSNQ and SDMT scores. This indicates that both fatigue and depression are strongly related to the degree of cognitive impairment measured by the MACFIMS.

SPSS output for these additional logistic regressions can be found in Appendix AD.

Summary

Combined scores for the MSNQ-I and SDMT produce ROC curves whose area under the curve approaches one in comparison to ROC curves produced for individual tests. Clinical use of this combined score is preferable to individual use of the MSNQ-I and SDMT.

For the MSNQ-P and SDMT the combined score was more accurate than the SDMT alone but in the case of the MSNQ-P the individual test produced a more accurate result. In this instance the combined use is preferable to use of the SDMT alone but the MSNQ-P test is superior when used alone rather than in combination with the SDMT.

Scores on both versions of the MSNQ were found to significantly correlate to scores on measures of both depression and fatigue. Additionally, post-hoc analysis of the FIS-Cognitive scores suggests it may be an accurate way to identify true cognitive impairment status for MS sufferers.

Discussion

Study aims, method and findings

The current study sought to establish validity for the joint use of the Multiple Sclerosis Neuropsychological Questionnaire (MSNQ) and Symbol Digit Modalities Test (SDMT) tests as screening tools for cognitive problems in people with Multiple Sclerosis (MS). To achieve this, a random sample was taken from participants recruited at MS clinics and MS drop-in facilities. Participants completed the MSNQ-Patient version (MSNQ-P) and SDMT tests and their informants (when available) completed the MSNQ-Informant test (MSNQ-I). Those randomly selected also completed the Minimal Assessment of Cognitive Function In Multiple Sclerosis (MACFIMS) cognitive test battery to establish a gold standard criterion of whether they were cognitively impaired or not. Using the method of McIntosh and Pepe (2002), logistic regression was carried out to establish a model for generating a combined score from both tests. This score was assessed using ROC analysis to establish sensitivity, specificity and accuracy for combined scores, which were compared with values derived from the ROC analysis carried out on the individual test results. Additional analysis was carried out to investigate the relationship between the MSNQ scores and measures of fatigue and low mood.

A total of 26 participants completed the MACFIMS battery of which 42.3% were classified as Impaired. There were no significant differences between the Impaired and Non-Impaired groups in terms of age, gender, ethnicity, MS subtype, years of education or years since diagnosis. ROC analysis on individual test scores only produced curves which differed from what would be expected by chance in the case of

the MSNQ-P. Using logistic regression, a formula for deriving a combined score for each MSNQ questionnaire in association with the SDMT was produced. However, the only statistically significant coefficient derived was that for the MSNQ-P.

Based on the results shown in Table 8 above, the levels of accuracy, sensitivity and specificity of the SDMT are improved by its use with the MSNQ-P to provide a combined score. Combined use of the SDMT with the MSNQ-I also improves accuracy and sensitivity compared to its use alone but there is not an improvement in specificity.

Production of a combined score for the MSNQ-P in combination with the SDMT does not improve levels of sensitivity, specificity or accuracy compared to use of the MSNQ-P alone. However, comparing use of the MSNQ-I alone to when it is combined with the SDMT superior accuracy and sensitivity are seen for the combined scores but specificity is not improved.

Based upon these findings it is recommended that the MSNQ-I be used in combination with the SDMT since this provides the greatest level of accuracy. If a single test is to be used then the MSNQ-P is recommended. This finding has great clinical significance since use of the combined scores for the MSNQ-I and SDMT will result in more appropriate referrals for neuropsychologists. Staff in MS clinics will also be able to use screening more effectively to determine which of their clients will most benefit from further testing, support and advice on rehabilitative strategies. MS patients will be able to obtain the additional help they require more quickly and be better enabled to know whether cognitive difficulties they perceive themselves to have are genuine impairments or not.

Based on previous findings that the MSNQ-P measure tends to be confounded by also measuring low mood, while the MSNQ-I does not. Correlational analyses were

also carried out on these scores in relation to participant scores on the BDI-FS. Both versions of the MSNQ were found to significantly correlate with depression scores. Since fatigue has also been associated with cognitive difficulties in MS, this analysis was also carried out for scores on the subscales of the FIS in relation to MSNQ scores. Both the patient and informant versions were found to have highly significant correlations to all three FIS subscales. The cognitive subscale of the FIS produced a ROC curve with superior accuracy to all of the individual and combined MSNQ and SDMT scores. A similar analysis on BDI-FS scores did not produce a statistically significant result.

Logistic regression that included BDI-FS and FIS scores alongside the screening test scores indicated that both depression and fatigue were strongly related to the degree of cognitive impairment measured by the MACFIMS. Values of Exp(B) from these analyses for depression and fatigue scores were equivalent or greater than those for MSNQ and SDMT scores suggesting that they make equal contributions to the model. It should be noted, however, that the B-values generated were not statistically significant.

Application of Results

Due to the underpowered sample, application of the results has to be made carefully, although trends can be identified which could lead to possibilities for future research. The following conclusions are therefore offered tentatively with acknowledgement of the underpowered nature of the study. Statistical power is required to ensure that results are not reported as non-significant when they may have been. This is of importance regarding the results of the logistic regressions. Although the only significant result obtained was for the MSNQ-P coefficient, this was indicative of a

possible trend towards significance as the sample approached power (the sample for the MSNQ-I regression was smaller and did not produce any significant results for coefficients). If this is correct, this is also suggestive of validity since, if the screening tools were not reliably relating to impairment, the model would not be able to be fitted at all. We can therefore conclude that the study provides possible evidence of validity for the two tools and their combined scores.

Further evidence of possible validity for the two tools and their combined scores is provided by the ROC analysis. A prerequisite for using the regression to optimise the ROC curve is that the scores actually relate to the disease status being identified in some way. The improvement in area under the curve, sensitivity and specificity values when the regression model is fitted provides evidence suggestive of the validity of the tests. The probability levels for these current results indicated it was unlikely to have occurred by chance and compared to the p-values for ROC analysis on individual tests, there is an increase in the statistical significance of the results which is again indicative of valid tests. The study demonstrated that use of the McIntosh and Pepe (2002) method for optimising the results of more than one screening test is practically possible with neuropsychological screening. There is sufficient reason to suggest that future research could select the MSNQ and SDMT alongside other screening tests and aim to combine them in a variety of combinations to see whether any could produce a very high level of accuracy or particularly strong sensitivity and specificity (perhaps seeking to achieve values for any of these of above 0.900). Of course, practical consideration could be given to how findings of such studies could be used in clinics. Since the screening tools investigated in the literature have already been selected or designed for their brevity and ease of use, the administration and marking of tests individually would not be problematic. Deriving a combined score would be best achieved in a clinical setting by

provision of a basic Excel spreadsheet that could use the coefficients derived from a study such as this one to allow practitioners to input test scores and easily obtain a combined score or even a predicted impairment status. Such a database could be made available for download from a website and then be used in MS clinics globally.

Questions about the validity of the MSNQ have been raised previously in the literature, with the patient version found to correlate with depression scores in many studies; and this was also found by the present research. However, a correlation has not previously been found for the Informant questionnaire. The findings of the current study indicate further research with larger samples should be carried out to establish whether the Informant questionnaire is indeed correlated to measures of low mood in the patient. Given that this result is contrary to previous findings it is only appropriate to be cautious about drawing final conclusions on this matter. Given the established relationship of fatigue to cognitive impairment, the correlations between MSNQ and FIS domains were unsurprising but also raise questions about whether the MSNQ conflates other associated problems with impairments or not. To be of value, the MSNQ-P should be a more 'pure' measure than this in order to avoid unnecessary referrals on to neuropsychology services.

The results of this study are important for a number of reasons. As discussed above they mean that nurses will be better able to make accurate decisions about which of their clients should be referred to neuropsychologists. Services will be better equipped to respond to the needs of their clients meaning they are even more in line with National Institute for Health and Clinical Excellence (NICE) guidance for how services should be provided (NICE, 2003). An additional benefit of having reliable and valid short screening tools is the ability to monitor MS patients' levels of cognitive impairment over time. Such tools could be administered at every clinic appointment for

this purpose with little intrusion into the limited time available to both clinician and client. Use of screening tools will result in improved services for people with MS.

Strengths and Limitations of the study

The current study was limited by not having achieved a large enough sample for an adequate logistic regression. Peduzzi, Concato, Kemper, Holford and Feinstein (1996) have established that a minimum sample size for meaningful results from a logistic regression requires that the number of cases under investigation in the dependent variable should be a minimum of 10 times the number of independent variables under analysis. Thus, a minimum of 20 Impaired participants was needed since each regression was carried out on only two variables. For the MSNQ-P analysis, a total of 11 Impaired participants were obtained, while for the MSNQ-I analysis, only 9 were identified. For the current study, there were around five Impaired participants per variable assessed, which would explain why results of the logistic regression on all but one variable (MSNQ-P) were not statistically significant. However, the emergence of a significant result on the MSNQ-P variable suggests that, had an adequately powered sample been achieved, the coefficients could have reached significance. Peduzzi et al. (1996) used multiple simulations to establish the distributions of coefficients from logistic regressions at different ratios of cases to variables. As this ratio decreases, the distribution becomes more flattened and dispersed and confidence intervals of the coefficients become wider. This is consistent with the results obtained by the current study where the standard errors on the coefficients were large, relative to the coefficients themselves.

Despite the underpowered sample for the logistic regression the optimisation process has produced superior levels of accuracy, sensitivity and specificity in the majority of the combined analyses compared to individual tests. However, caution also needs to be exercised with interpretation of these results since samples are also underpowered for adequate ROC analysis. Using the method of Buderer (1996), a sample size of 70 participants would have been needed for adequate power. This would explain the non-significance of results on some of the individual test ROC analyses. However, the superior levels of accuracy, sensitivity and specificity for combined scores suggests that there is a trend towards improved scores due to the logistic regression method of McIntosh and Pepe (2002). Nonetheless, further research is needed to be able to rule out the possibility of a Type I error in this instance.

The lack of adequate power for the current study can be attributed to a combination of factors. Firstly the study had to be completed to a tight deadline which, combined with an unduly lengthy Trust approval process meant that it was not possible to achieve power and still complete the study report in the available time. The process of completing Stage Two testing was time consuming because of missed appointments by the participants who were selected. Of 33 participants selected for this stage of the study, only 26 returned for testing. Over one fifth of those to be tested either opted out of the study or were unable to attend. Attempts were made to contact those who did not attend to discover why this was, but responses were not received. For many who did attend Stage Two testing, obtaining time off work was not always straightforward, as employers felt the research was not a medical appointment. While this could introduce some inadvertent bias into the sample, if many people did not attend for this reason, it is not clear that results would be adversely affected. People may be unemployed due to

physical disablement of MS which is not consistently found by research to correlate to cognitive impairments.

The current research benefits from a more rigorous methodology than many of the published studies and is particularly strong regarding the use of random selection between Stages One and Two. This ensured the sample used was not biased towards clients who attended the clinic earlier in the recruitment period or those who were more keen to take part than others. However, this random selection element was problematic for the current study in terms of speed at which data could be collected. Since three participants had to be recruited for every participant who was tested, this prevented further selection to Stage Two until enough participants had become involved at Stage One. Future studies would benefit from more generous randomisation criteria such that one in every two Stage One participants would be selected. This would also prevent non-attendance being as significant a set-back as it was in the present study, since there would be a more steady flow of participants coming to testing from Stage One.

It was hoped to also introduce blinding of the experimenter to screening test results when carrying out the MACFIMS testing battery with participants. It had originally been proposed that staff at MS clinics would carry out the consent procedure and screening tests but this was not logistically possible for them, so the researcher was required to carry out this aspect of the study. However, an attempt at blinding was made by not marking screening tests until after completion of the MACFIMS battery.

Use of an established and validated battery of tests selected by a panel of experts explicitly for use in the MS population greatly increased the quality of the methodology. However, it is of value to note that the norms for each of these tests have been generated separately in disparate populations. There is debate within

neuropsychology regarding the value of using regression based norms to interpret test performance. Such norms allow a predicted performance to be established based on a variety of demographic variables such as age, gender and level of education against which actual performance can be compared. At the time of planning the current study and applying for ethical approval, the researcher was unaware of a study by Parmenter, Testa, Schretlen, Weinstock-Guttmann and Benedict (2010) which derived regression-based norms for the MACFIMS battery. Parmenter et al. (2010) noted that the use of these norms identified more of their MS participants as 'impaired' than standard manualised norms. They concluded that the norms provided a valid method for establishing levels of impairment when using the MACFIMS. Use of these norms in a future study may provide a suitable sample for logistic regression more quickly and would also prevent those at the borderlines of age groups in the manualised norms from being unfairly judged against the performance of those who may be 10 or 15 years younger or older. It would also reduce a possible bias regarding differences between 'impaired' and 'non-impaired' groups on level of education. While the difference between groups was not statistically significant, it was close to being so ($p=0.055$). Regression-based norms would take this into account and reduce any potential impact upon the results.

The final area to consider with regard to strengths and weaknesses of the current study is the involvement of informants in the research. Whenever the MSNQ has been assessed in the literature, it is usually stated that participants and informants completed their versions of the questionnaire at the same time at the MS clinic. However, in the current study, around 50% of those attending a clinic arrived without anyone who could act as an informant. A decision was made to allow recruitment of these people to avoid biasing the sample and asking participants to bring an informant to later testing if

selected. Those who arrived with an informant to clinic are more likely to have severe physical or cognitive problems and as such, may not be representative of the wider MS population. Given the findings of stability in MSNQ scores across time by Morrow et al. (2010), the later completion of the MSNQ-I by some informants should not be problematic in terms of whether such scores are reliable but it did reduce the overall quantity of informant data that could be analysed as some participants could not provide an informant at later testing.

Clinical Implications

Based on the values of Exp(B) obtained from the logistic regressions incorporating scores on the BDI-FS and FIS domains, it was clear that when conducting testing using the MACFIMS battery, it is important for neuropsychologists to also conduct testing of mood and fatigue levels in order to avoid confounding their contribution to impairment scores with the results of disease action. The current study confirmed the recommendation of the consensus panel that constructed the battery for such testing to be included (Benedict et al., 2002).

For the MS population, it is preferable to reduce as far as possible the number of false negatives produced by a test. While a false positive would possibly produce undue distress in a patient, it would produce a referral for more comprehensive testing, which would allow fears to be allayed and the true situation to be established. In the case of a false negative, however, further testing is unlikely to occur, meaning that those with genuine impairments would be denied access to a service that would be of use to them. As a result, considerations of where to set the cut-off point for impairment should be made such as to maximise sensitivity of the test in preference to specificity in any trade off between the two.

Future Research

The finding of superior values of accuracy, sensitivity and specificity for the cognitive domain score of the Fatigue Impact Scale was of great interest. Further investigation should be carried out to discover whether this particular test could be used for cognitive screening. Such research should investigate the qualities found by ROC analysis on the test scores in a larger sample than that presented here. The FIS has many of the qualities required of a good screening test – it requires little training to administer and to score and can be completed quickly in a clinical environment.

It would be of great value to continue the data collection carried out thus far into the future in order to achieve a sample with sufficient power to establish statistically significant B-values in the logistic regression model. This would be expected to further optimise the results obtained from ROC analysis, bringing the area under the curve even closer to 1.000. If, as suggested above, further research also investigates the combining of other screening tests with each other, a good starting point for such research would be to examine performance on a short test of attention, memory or executive function in association with a test concerning processing speed such as the SDMT.

Conclusions

The current study provided evidence in principle of the value of combining scores from neuropsychological screening tests. Although much of the statistical analysis is underpowered the current results were strongly suggestive of the usefulness and applicability of McIntosh and Pepe's (2002) method in screening for cognitive impairments in MS patients. Future research is needed to verify the validity of these provisional findings in an adequately powered sample.

References

Akbar, N., Honarmand, K., Kou, N., Levine, B., Rector, N. & Feinstein, A. (2010) Validity of an internet version of the multiple sclerosis neuropsychological questionnaire. *Multiple Sclerosis*, 16 (12), 1500-1506.

Amato, M.P., Portaccio, E., Goretti, B., Zipoli, V., Hakiki, B., Giannini, M., Pasto, L. & Razzolini, L. (2010) Cognitive impairment in early stages of multiple sclerosis. *Neurological Science*, 31 (Suppl. 2), S211-S214.

Arrondo, G., Alegre, M., Sepulcre, J., Iriate, J., Artieda, J., & Villoslada, P. (2009) Abnormalities in brain synchronisation are correlated with cognitive impairment in multiple sclerosis. *Multiple Sclerosis*, 15, 509-516.

Banati, M., Sandor, J., Mike, A., Illes, E., Bors, L., Feldmann, A., Herold, R. and Illes, Z. (2010) Social cognition and theory of mind in patients with relapsing-remitting multiple sclerosis. *European Journal of Neurology*, 17, 426-433.

Barak, Y. and Achiron, A. (2006) Cognitive fatigue in multiple sclerosis: Findings from a two-wave screening project. *Journal of the Neurological Sciences*, 245, 73-76.

Basso, M.R., Beason-Hazen, S., Lynn, J., Rammohan, K. and Bornstein, R.A. (1996) Screening for cognitive dysfunction in multiple sclerosis. *Archives of Neurology*, 53, 980-984.

Basso, M.J., Candilis, P.J., Johnson, J., Ghormley, C., Combs, D.R. & Ward, T. (2010) Capacity to make medical treatment decisions in multiple sclerosis: A potentially remediable effect. *Journal of Clinical and Experimental Neuropsychology*, 32 (10), 1050-1061.

Bate, A.J., Matthias, J.L. & Crawford, J.R. (2001) Performance on the Test of Everyday Attention and standard tests of attention following severe traumatic brain injury. *The Clinical Neuropsychologist*, 15 (3), 405-422.

Beal, C.C., Stuifbergen, A.K. & Brown, A. (2007) Depression in multiple sclerosis: a longitudinal analysis. *Archives of Psychiatric Nursing*, 21 (4), 181-191.

Beatty, W.W. & Monson, N. (1994) Picture and motor sequencing in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, 16, 165-172.

Beatty, W.W., Paul, R.H., Wilbanks, S.L., Hames, K.A., Blanco, C.R. and Goodkin, D.E. (1995a) Identifying multiple sclerosis patients with mild or global cognitive impairment using the screening examination for cognitive impairment (SEFCI). *Neurology*, 45, 718-723.

Beatty, W.W., Hames, K.A., Blanco, C.R., Paul, R.H. & Wilbanks, S.L. (1995b) Verbal abstraction deficit in multiple sclerosis. *Neuropsychology*, 9 (2), 198-205.

Beatty, W.W. & Monson, N. (1996) Problem solving by patients with multiple sclerosis: Comparison of performance on the Wisconsin and California Card Sorting Test. *Journal of the International Neuropsychological Society*, 2, 134-140.

Beck, A.T., Steer, R.A., Brown, G.K. (2000) *BDI-Fast Screen for Medical Patients: Manual*, San Antionio, TX: Psychological Corporation.

Bellmann-Strobl, J., Wuerfel, J., Aktas, O., Dorr, J., Wernecke, K.D., Zipp, F. & Paul, F. (2009) Poor PASAT Performance Correlates with MRI Contrast Enhancement in Multiple Sclerosis. *Neurology*, 73, 1624-1627.

Benedict, R.H.B., Schretlen, D., Groninger, L., Dobraski, M. & Shpritz, B. (1996) Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability and validity. *Psychological Assessment,* 8, 145-153.

Benedict, R.H.B (1997) *Brief Visuospatial Memory Test – Revised: Professional Manual, Lutz, FL:* Psychological Assessment Resources.

Benedict, R.H.B., Priore, R.L., Miller, C., Munschauer, F. and Jacobs, L. (2001) Personality disorder in multiple sclerosis correlates with cognitive impairment. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 13, 70-76.

Benedict, R.H.B., Fischer, J.S., Archibald, C.J., Arnett, P.A., Beatty, W.W., Bobholz, J., Chelune, G.J., Fisk, J.D., Landgdon, D.A., Caruso, L., Foley, F., LaRocca, N.G., Vowels, L., Weinstein, A., DeLuca, J., Rao, S.M. and Munschauer, F. (2002) Minimal Neuropsychological Assessment of MS Patients: A Consensus Approach, *The Clinical Neuropsychologist*, 16, 381-397.

Benedict, R.H.B., Munschauer, F., Linn, R., Miller, C., Murphy, E., Foley, F., and Jacobs, L. (2003a) Screening for Multiple Sclerosis Cognitive Impairment Using a Self-Administered 15-Item Questionnaire. *Multiple Sclerosis*, 9, 95-101.

Benedict, R.H.B., Fishman, I., McClellan, M.M., Bakshi, R. & Weinstock-Guttman, B. (2003b) Validity of the Beck Depression Inventory – Fast Screen in multiple sclerosis. *Multiple Sclerosis*, 9, 393-396.

Benedict, R.H.B., Cox, D., Thompson, L.L., Foley, F., Weinstock-Guttmann, B. & Munschauer, F. (2004) Reliable Screening for Neuropsychological Impairment in Multiple Sclerosis. *Multiple Sclerosis*, 10, 675-678.

Benedict, R.H.B., Cookfair, D., Gavett, R., Gunther, M., Munschauer, F., Garg, N. & Weinstock-Guttman, B. (2006) Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, 12, 549-558.

Benedict, R.H.B., Duquin, J.A., Jurgensen, S., Rudick, R.A., Feitcher, J., Munschauer, F.E., Panzara, M.A. & Weinstock-Guttman, B. (2008) Repeated assessment of neuropsychological deficits in multiple sclerosis using the Symbol Digit Modalities Test and the MS Neuropsychological Screening Questionnaire. *Multiple Sclerosis*, 14, 940-946.

Benito-Leon, J. (2011) Multiple Sclerosis: Is prevalence rising and if so why? *Neuroepidemiology*, 37, 236-237.

Benton, A.L., Hamsher, K.D., Varney, N.R., and Spreen, O. (1983) *Judgement of line orientation,* New York: Oxford University Press.

Benton, A.L. and Hamsher, K. (1989) *Multilingual Aphasia Examination*, Iowa City, IA: AJA Associates.

Benton, A.L., Sivan, A.B., Hamsher, K. deS., Varney, N.R. & Spreen, O. (1994) C*ontributions to Neuropsychological Assessment (2nd Ed.),* Orlando, Florida: Psychological Assessment Resources.

British Psychological Society (2010) *Code of Human Research Ethics,* Leicester, UK: British Psychological Society.

Buderer, N.M.F. (1996) Statistical Methodology: I. Incorporating the Prevalence of Disease into the Sample Size Calculation for Sensitivity and Specificity. *Academic Emergency Medicine*, 3, 895-900.

Calabrese, M., Rinaldi, F., Mattisi, I., Grossi, P., Favaretto, A., Atzori, M., Bernardi, V., Barachino, L., Romualdi, C., Rinaldi, L., Perini, P. & Gallo, P. (2010) Widespread cortical thinning characterizes patients with MS with mild cognitive impairment. *Neurology*, 74, 321-328.

Chamelian, L., Bocti, C., Gao, F-Q., Black, S.E. & Feinstein, A. (2005) Detecting cognitive dysfunction in multiple sclerosis with a magnetic resonance imaging rating scale: a pilot study. *CNS Spectrums*, 10 (5), 394-401.

Crawford, J.R., Obansawin, M.C. & Allan, K.M. (1998) PASAT and components of WAIS-R performance: convergent and discriminant validity. *Neuropsychological Rehabilitation*, 8 (3), 272-273.

Delis, D.C., Kramer, J.H., Kaplan, E. & Ober, B.A. (2000) *California Verbal Learning Test Manual Second Edition: Adult Version*, San Antionio, TX: Psychological Corporation.

Delis, D.C., Kaplan, E., Kramer, J.H. (2001a) *Delis-Kaplan Executive Function System (DKEFS): examiner's manual*, San Antonio, TX: The Psychological Corporation.

Delis, D.C., Kaplan, E., Kramer, J.H. (2001b) *Delis-Kaplan Executive Function System (DKEFS):Technical Manual*, San Antonio, TX: The Psychological Corporation.

Deloire, M.S.A., Bonnet, M.C., Salort, E., Arimone, Y., Boudineau, M., Petry, K.G. & Brochet, B. (2006) How to detect cognitive dysfunction at early stages of multiple sclerosis? *Multiple Sclerosis*, 12, 445-452.

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.

Engel, C., Greim, B. and Zettl, U.K. (2007) Diagnostics of Cognitive Dysfunctions in Multiple Sclerosis. *Journal of Neurology*, 254 (Suppl. 2), II/30-II/34.

Fischer, J.S., Jak, A.J., Kniker, J.E., Rudick, R.A. & Cutter, G. (2001) *Multiple Sclerosis Functional Composite (MSFC): Administration and Scoring Manual*, New York, NY: National Multiple Sclerosis Society.

Fisk, J.D., Ritvo, P.G., Ross, L., Haase, D.A., Marrie, T.J., Schlech, W.F. (1994). Measuring the Functional Impact of Fatigue: Initial Validation of the Fatigue Impact Scale. *Clinical Infectious Diseases*, 18 (Suppl.1), S79-83.

Galeazzi, G.M., Ferrari, S., Giaroli, G., Mackinnon, A., Merelli, E., Motti, L. & Rigatelli, M. (2005) Psychiatric disorders and depression in multiple sclerosis outpatients: impact of disability and interferon beta therapy. *Neurological Science*, 26, 255-262.

Glanz, B.I., Healy, B.C., Rintell, D.J., Jaffin, S.K., Bakshi, R. & Weiner, H.L. (2010) The association between cognitive impairment and quality of life in patients with early multiple sclerosis. *Journal of Neurological Sciences*, 290, 75-79.

Gronwall, D.M. (1977) Paced Auditory Serial Addition Task: A Measure of Recovery from Concussion. *Perceptual and Motor Skills*, 44(2), 367-373.

Haase, C.G., Lienemann, M. & Faustmann, P.M. (2008) Neuropsychological deficits but not coping strategies are related to physical disability in multiple sclerosis. *European Archives of Psychiatry and Clinical Neuroscience*, 258, 35-39.

Hedstrom, A.K., Akerstedt, T., Hillert, J., Olsson, T. and Alfredson, L. (2011) Shift work at young age is associated with increased risk for multiple sclerosis. *Annals of Neurology*, 70, 733-741.

Julian, L. (2011) Cognitive functioning in multiple sclerosis. *Neurologic Clinics*, 29 (2), 507-525.

Julian, L., Vella, L., Vollmer, T., Hadjimichael, O. & Mohr, D.C. (2008) Employment in multiple sclerosis: exiting and re-rentering the workforce. *Journal of Neurology*, 255, 1354-1360.

Kalmar, J.H., Gaudino, E.A., Moore, N.B., Halper, J. & DeLuca, J. (2008) The relationship between cognitive deficits and everyday functional activities in multiple sclerosis. *Neuropsychology*, 22 (4), 442-449.

Kurtzke, J.F. (1983) Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*, 33 (11), 1444-1452.

Lacy, M.A., Gore, P.A., Pliskin, N.H., Henry, G.K., Heilbroner, R.L. & Hamer, D.P. (1996) Verbal fluency task equivalence. *Clinical Neuropsychologist*, 10 (3), 305-308.

Mathiowetz, V. (2003) Test-retest reliability and convergent validity of the Fatigue Impact Scale for persons with Multiple Sclerosis. *American Journal of Occupational Therapy*, 57, 389-395.

McCaffrey, R.J., Cousins, J.P., Westervelt, H.J., Martnowicz, M., Remick, S.C., Szebenyi, S., Wagle, W.A., Bottomley, P.A., Hardy, C.J., & Haase, R.F. (1995) Practice Effect with the NIMH AIDS Abbreviated Neuropsychological Battery. *Archives of Clinical Neuropsychology*, 10, 241-250.

McIntosh, M.W. & Pepe, M.S. (2002) Combining Several Screening Tests: Optimality of the Risk Score. *Biometrics*, 58, 657-664.

Mohr, D.C., Hart, S.L., Julian, L. & Tasch, E.S. (2007) Screening for depressionamong patients with multiple sclerosis: two questions may be enough. *Multiple Sclerosis*, 13, 215-219.

Morgan, S.F. & Wheelock, J. (1992) Digit Symbol and Symbol Digit Modalities Tests: Are they directly interchangeable? *Neuropsychology*, 4 (6), 327-330.

Morrow, S.A., O'Connor, P.W., Polman, C.H., Goodman, A.D., Kappos, L., Lublin, F.D., Rudick, R.A., Jurgensen, S., Paes, D., Forrestal, F. & Benedict, R.H.B. (2010) Evaluation of the symbol digit modalities test (SDMT) and MS neuropsychological screening questionnaire (MSNQ) in natalizumab-treated MS patients over 48 weeks. *Multiple Sclerosis*, 16 (11), 1385-1392.

National Institute for Health and Clinical Excellence (2003) *Multiple Sclerosis: Mangement of Multiple Sclerosis in Primary and Secondary Care*, London, UK: National Institute for Clinical Excellence.

Norris, M.P., Blankenship-Reuter, L., Snow-Turek, A.L. & Finch, J. (1995) Influence of depression on verbal fluency performance. *Aging and Cognition*, 2 (3), 206-215.

O'Brien, A., Gaudino-Goering, E., Shawaryn, M., Komaroff, E., Moore, N.B. & DeLuca, J. (2007) Relationship of the multiple sclerosis neuropsychological questionnaire (MSNQ) to functional, emotional, and neuropsychological outcomes. *Archives of Clinical Neuropsychology*, 22, 933-948.

Olazaran, J., Cruz, I., Benito-Leon, J., Morales, J.M., Duque, P., Rivera-Navarro, J. (2009) Cognitive dysfunction in multiple sclerosis: Methods and prevalence from the GEDMA study. *European Neurology*, 61 (2), 87-93.

Ouellet, J., Scherzer, P.B., Rouleau, I., Metras, P., Bertrand-Gauvin, C., Njerroud, N., Boisseau, E. and Duquette, P. (2010) Assessment of social cognition in patients with multiple sclerosis. *Journal of the International Neuropsychological Society*, 16, 287-296.

Parmenter, B.A., Weinstock-Guttmann, B., Garg, N., Munschauer, F. and Benedict, R.H.B. (2007) Screening for Cognitive Impairment in Multiple Sclerosis Using the Symbol Digit Modalities Test. *Multiple Sclerosis*, 13, 52-57.

Parmenter, B.A., Testa, S.M., Schretlen, D.J., Weinstock-Guttman, B. & Benedict, R.H.B. (2010) The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, 16 (1), 6-16.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R. & Feinstein, A.R. (1996) A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis. *Journal of Clinical Epidemiology*, 49 (12), 1373-1379.

Pintea, S. and Moldovan, R. (2009) The Receiver Operating Characteristic (ROC) Analysis: Fundamentals and Applications in Clinical Psychology. *Journal of Cognitive and Behavioural Psychotherapies*, 9 (1), 49-66.

Prakash, R.S., Snook, E.M., Lewis, J.M., Motl, R.W. & Kramer, A.F. (2008) Cognitive impairments in relapsing-remitting multiple sclerosis: a meta-analysis. *Multiple Sclerosis*, 14, 1250-1261.

Pugliatti, M., Sotgiu, S. and Rosati, G. (2002) The worldwide prevalence of multiple sclerosis. *Clinical Neurology and Neurosurgery*, 104, 182-191.

Qualls, C.E., Bliwise, N.G. & Stringer, A.Y. (2000) Short Forms of the Benton Judgement of Line Orientation Test: Development and Psychometric Properties. *Archives of Clinical Neuropsychology*, 15, 159-163.

Rao, S.M., Leo, G.J., Bernardin, L. and Unverzagt, F. (1991a). Cognitive dysfunction in multiple sclerosis: I. Frequency, patterns and prediction. *Neurology*, 41, 685-691.

Rao, S.M., Leo, G.J., Ellington, L., Nauertz, T., Bernardin, L. and Unverzagt, F. (1991b). Cognitive dysfunction in multiple sclerosis: II. Impact on employment and social functioning. *Neurology*, 41, 692-696.

Rao, S. (1995) Neuropsychology of multiple sclerosis. *Current Opinion in Neurology*, 8, 216-220.

Ruff, R.M., Light, R.H., Parker, S.B. & Levin, H.S. (1996) Benton controlled oral word association test: reliability and updated norms. *Archives of Clinical Neuropsychology*, 11 (4), 329-338.

Ruggieri, R.M., Palermo, R., Vitello, G., Gennuso, M., Settipani, N. & Piccoli, F. (2003) Cognitive impairment in patients suffering from relapsing-remitting multiple sclerosis with EDSS ≤ 3.5. *Acta Neurologica Scandinavica*, 108, 323-326.

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, 176-193.

Santon, A., Cristobal, E., Aparicio, M., Royuela, A., Villar, L.M. and Alvarez-Cermeno, J.C. (2011) High frequency of co-infection by Epstein-Barr virus types 1 and 2 in patients with multiple sclerosis. *Multiple Sclerosis Journal*, 17 (11), 1295-1300.

Schultheis, M.T., Garay, E. & DeLuca, J. (2001) The influence of cognitive impairment on driving performance in multiple sclerosis. *Neurology*, 56, 1089-1094.

Simpson, S., Blizzard, L., Otahal, P., Van der Mei, I. and Taylor, B. (2011) Latitude is significantly associated with the prevalence of multiple sclerosis: A meta-analysis. *Journal of Neurology, Neurosurgery and Psychiatry*, 82, 1132-1141.

Smith, A.(1982) *Symbol Digit Modalities Test*. Los Angeles, California: Western Psychological Services.

Smith, A.(1991) *Symbol Digit Modalities Test*. Los Angeles, California: Western Psychological Services.

Sperling, R.A., Guttman, C.R., Hohol, M.J., Warfield, S.K., Jakab, M., Parente, M., Diamond, E.L., Kikinis, R., Jolesz, F.A. & Weiner, H.L. (2001) Regional magnetic resonance imaging lesion burden and cognitive function in multiple sclerosis: A longitudinal study. *Archives of Neurology*, 58, 115-121.

Thornton, A.E., Raz, N. & Tucke, K.A. (2002) Memory in multiple sclerosis: contextual encoding deficits. *Journal of the International Neuropsychological Society*, 8, 395-409.

Tombaugh, T.N. (2006) A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). *Archives of Clinical Neuropsychology*, 21, 53-76.

Trahan, D.E. (1998) Judgment of Line Orientation in patients with unilateral cerebrovascular lesions. *Assessment*, 5, 227-235.

Vanderploeg, R.D., LaLone, L.V., Greblo, P., & Schinka, J.A. (1997) Odd-Even short forms of the Judgment of Line Orientation Test. *Applied Neuropsychology*, 4, 244-246.

Vanotti, S., Benedict, R.H.B., Acion, L., Caceres, F. & VANEM Workgroup (2009) Validation of the multiple sclerosis neuropsychological questionnaire in Argentina. *Multiple Sclerosis*, 15, 244-250.

Winegarden, B.J., Yates, B.L., Moses, J.A., Benton, A.L. & Faustman, W.O. (1998) Development of an optimally reliable short form for Judgment of Line Orientation. *The Clinical Neuropsychologist*, 12, 311-314.

Woodward, J.L., Benedict, R.H.B., Salthouse, T.A., Toth, J.P., Zgaljardic, D.J. & Hancock, H.E. (1996) Normative data for equivalent, parallel forms of the Judgment of

Line Orientation Test. *Journal of Clinical and Experimental Neuropsychology*, 20, 457-462.

Younes, M., Hill, J., Quinless, J., Kilduff, M., Peng, B., Cook, S.D. & Cadavid, D. (2007) Internet-Based Cognitive Testing in Multiple Sclerosis. *Multiple Sclerosis*, 13, 1011-1019.

Ziemsen, T. (2009) Multiple sclerosis beyond EDSS: depression and fatigue. *Journal of the Neurological Sciences*, 277 (S1), S37-S41.

Zipoli, V., Goretti, B., Hakiki, B., Siracusa, G., Sorbi, S., Portaccio, E. & Amato, M.P. (2010) Cognitive impairment predicts conversion to multiple sclerosis in clinically isolated syndromes. *Multiple Sclerosis*, 16 (1), 62-67.

Critical Appraisal

Introduction

The research process has been both a challenge and a learning experience and this document sets out the aspects of this experience from initial conception of the research idea through to completion of the thesis. I have structured this account chronologically, allowing the development of the project to be seen as a process and highlighting reflections and learning points as they occurred in the course of the research.

Initial Research Ideas (Optimistic)

Upon starting the Doctorate in Clinical Psychology I had already given consideration to areas that would be of interest to me for my research project. The thinking behind this had been that, since I would need to sustain interest in the project for some two to three years, it would be important it was something I had some level of interest or passion for. A number of broad areas were possibilities at this stage including research into staff experiences of providing mental health services and investigation of the experience of hearing voices. However, Multiple Sclerosis was always of most interest for personal reasons. In early 2008 I had received an unexpected phone call from my brother telling me he had just been diagnosed with MS. Up to this point I had, like most people, known little about MS but my brother's diagnosis prompted me to learn more about this illness and generated a desire to further the research in the area. When I found out at a research conference that there was a neuropsychologist who had a project regarding MS that had previously gone through the ethical process successfully but not yet been carried out, this seemed ideal. I organised to meet the neuropsychologist concerned and was convinced to take the project on since it was something about which I found myself immediately enthusiastic.

Additionally, the neuropsychologist described herself and the team in the department as 'all being a bit science geeky' which convinced me I would fit in with them well. I decided to carry out my first year literature review in the topic to discover more about it and felt that this was something that would sustain my interest for the duration of the project.

An initial research proposal was submitted to the University in June 2009 based on the work previously done on the project by Dr. Kirkland and my own literature review. An internal reviewer of the proposed study suggested that the sample size to obtain was possibly overly optimistic and some concerns were voiced about the use of one of the tests – the Paced Auditory Serial Addition Test – as often being experienced by patients as quite aversive. However, my research supervisor assured me he felt confident the sample was achievable and, given that the project had originally been designed to be carried out as part of an MSc project, I told myself that I was worrying unduly. Changes were made to the research protocol such that the PASAT would be administered last (allowing time for discussion of the experience with participants after completion) and introduced as sometimes being difficult for people. As with all psychometrics, the PASAT would not be administered if participants expressed they did not wish to complete it.

Ethics process (In Limbo)

My previous research experience had only required me to submit my research proposal to a University ethics committee so the prospect of having to gain approval from an NHS Research Ethics Committee and the local Research & Development Department at the hospital concerned seemed quite daunting at first. This was, however,

tempered by the fact that I knew the project had already gained approval and so I did not anticipate the process to be too problematic.

Since the proposed project was not identical to the originally approved project my supervisor and I decided to reference the earlier approval in our ethics applications but make clear there had been some slight alterations to the project. Completion of the IRAS form was approached as a long-term project as trainees from cohorts before me had told me the form was lengthy and complex to complete. As a result I was able to complete the form in what I considered manageable chunks and booked an appointment at a meeting of the Leicestershire, Northamptonshire & Rutland Research Ethics Committee 2 after submitting the IRAS application form on 31st January 2011. Despite an initial request to make slight amendments to the Consent Forms on 2nd Feb 2011, the meeting with the REC still took place on 17th February as had originally been booked and my experience of it was that, while naturally anxiety provoking, the meeting went well and the committee were genuinely interested in what I wished to do. This was a pleasant surprise as the REC committees had a reputation for being awkward and fussy. My experience of them was anything but and I was impressed that service users were represented on the committee alongside a range of health professionals. The committee were prompt in responding to me after the meeting and requested a total of four items of further information all of which were quite straightforward and non-problematic to provide. A positive ethical opinion was provided by the REC on 19th April 2011.

Invigorated by the positive experience of applying to the ethics committee, I moved on to gain approval for the study from the University Hospitals of Leicester NHS Trust Research and Development Office. This process was begun on 19th April 2011 when the study was registered in UHL's system. However, soon after this, I received an email from the R&D team telling me they could not continue with my

application until I submitted a substantial amendment to my research protocol. They raised a concern about the way in which information would be shared with participants if psychological impairments were found which were felt to put them at future risk. While my protocol said I would inform the person responsible for their care of this, it also said that responsibility for informing the participant at the time would lie with me as researcher. The R&D team felt this would not sit comfortably within the parameters of my proposed activity or my interaction with UHL patients. There was a learning point here that is now clear to me – that ethical expectations when working in research in a University are different from those in an NHS setting where there are lines of responsibility and accountability for care. The change seemed trivial for me to make and was perfectly acceptable to me but having to submit a substantial amendment to protocol held up the process of approval by around two months for the sake of what seemed to me a matter of the wording of the protocol. This was to be typical of my interactions with the R&D unit in that much of what they required was precisely worded documents in printed and digital formats with exactly correct version numbers and on the correct form of headed paper. While this was initially simple enough to provide, it became more complex as the department produced a number of minor changes. All of these required a new version number on the changed document which meant it had to be provided again in every format and also be given back to the REC to be approved as a minor amendment. The approval for the minor amendment then had to be provided in both printed and digital format to the R&D department. An additional complication came when, despite having provided the documents required, I was told on a few occasions that a copy had not been received. Some items had to be sent to the department up to three times before being accepted as having been seen by them and this occurred frequently despite my use of telephone calls to explain clearly the names

and details of files they required and when they were sent. However, with great patience and perseverance these aspects of the process were able to be negotiated and it can be understood why, for the purposes of ensuring participants are treated correctly, that specific wordings that make everything clear for them, are placed on consent forms and information sheets. Nonetheless I experienced this process as quite emotionally draining and frustrating. It felt as though I was left in limbo – feeling personally ready and eager to commence research but waiting on the full workings of a bureaucratic process that seemed to be taking longer than I thought necessary.

In early June 2011, I was asked by the UHL finance department to give a full breakdown in a 10 page Excel spreadsheet they provided to me of every costing that the study would produce for UHL itself. Upon contacting them to ask further about this I was also informed I would need to pay a fee of £730 to register my study with them. This was the first of many points where the requirements of the R&D process began to feel more than a little overwhelming for me. I had anticipated approval having been gained by this stage in the year and was now eager to commence data collection but it felt as though more and more requirements were being produced by a process that now appeared to have no end in sight. However, after successfully contacting and explaining my work with a project manager at UHL's finance department, I was told that given there would be no financial cost to UHL, the finance department could approve my study and that I would no longer have to pay the registration fee.

Given I had not met my first deadline for commencing data collection, I reappraised my goals for the study. It was now early July and there were only five more requirements from UHL R&D before I could gain their approval. The first of these two were that I was required to undergo training in order to be certified as knowledgeable about research governance practices and I had to be assessed as capable of gaining

informed consent from participants. Upon attempting to organise both of these, I was told that the people in the team who could carry out the consent assessment would be on annual leave such that it would not be possible to organise this for one month. After further negotiation, a hasty assessment was arranged and I was able to satisfy the R&D department of my ability to take proper informed consent. My training in research governance occurred the same week. This left me three matters to sort out prior to my flying away to San Francisco for my wedding and honeymoon at the end of July. I was required to have a member of UHL staff to take management responsibility for me while on their premises; a section of my form needed to be signed by a senior medic at Leicester Royal Infirmary; and a further minor amendment was needed to be approved by the REC. Various emails and phone calls were made such that by 22nd July I knew I could fly away and things would be sorted out on my return three weeks later. On my return, none had actually been resolved, so further chasing and discussions took place with email confirmation of management and REC approval arriving by the end of the week. The senior medic's secretary informed me that the necessary form had been signed and sent but R&D informed me it had not been received, so a trip to Leicester Royal Infirmary was made to collect a duplicate signed form which I delivered in person to the UHL R&D office in exchange for a receipt. After this final matter was resolved, I was finally given approval on 25th August 2011 for my study to commence. It was with immense relief that I exited what had become for me quite an alienating and frustrating experience.

Data Collection (Like Spinning Plates)

Having received permission to commence data collection, I tried to begin the research process as quickly as possible but encountered another hurdle immediately. During the ethics process, I had arranged with a research nurse based at the MS clinic I

would be recruiting from, for my information sheets and consent forms to be included in envelopes with the information she was sending regarding the studies she was running. However, upon meeting again with the research nurse, I was told this would no longer be possible as she felt this would be likely to harm recruitment to *her* study. Without sending information to potential participants prior to their appointments, I was unable to approach them in the clinics to discuss participation. A meeting between myself, the Consultant for the clinic and my supervisor was hastily arranged in the hope that administrative staff at the clinic would be able to send my forms with appointment letters, but it was made clear to us that this would not be feasible either. For a while it seemed the project was 'dead in the water' since I did not feel I would be allowed to access files of clients to send the information out myself. However, staff at the MS clinic suggested I could do this and so I contacted UHL's R&D department to discuss the issue. To my surprise they were happy for me to take the addresses of those with forthcoming appointments from the fronts of files in the clinic in order to send details out so long as no information was taken out of the department. As the only solution to the problem, I immediately went ahead with this, knowing it would have inevitable consequences for my time and budget. As well as gaining consent, completing and scoring all the psychometric tests and collating and analysing data, I would now need to spend a substantial period of time addressing and stuffing envelopes with study details. There was also now an additional cost to my study budget in postage for sending information to participants, which had not been previously accounted for. However, I felt I had no choice but to proceed given that it was now September 2011 and the study needed to be completed and handed in by April 2012.

This setback aside, the process of data collection was immensely enjoyable. I felt it a great honour to meet so many people with MS. Inevitably I got to hear many

people's stories of the experience of diagnosis and their attitudes to their illnesses. As a group, those with MS appeared to me to be very determined and resilient people. Prior to beginning data collection I had been told that people with an MS diagnosis are very keen to participate in research projects and this proved to be true. Recruitment to the study occurred at a pace that began to calm my concerns about obtaining the sample required in time. It was common for participants to describe their experience of diagnosis as being a process where for many years there were a series of abnormal physical or sensory experiences, which now made sense in light of their illness. Many people told me how they experienced a lack of understanding from others of their condition as MS is poorly understood and can have little impact if in a remission phase but be totally disabling when in a relapse. This experience of 'invisible disability' was even more marked for the cognitive losses people experienced, with these symptoms being even less evident to others than the physical ones.

It quickly became evident that despite my information sheets making clear that I would need someone to act as an informant, a very significant number of participants were not able to have an informant accompany them to the MS Clinic. This presented a problem for recruitment of informants, as time left to obtain a sample was short, so a decision was made that if potential participants did not have an informant at this stage then they would be explicitly asked to bring one along at the second stage testing if selected. This seemed the best compromise between obtaining the data required and still being able to recruit at a realistically achievable speed. However, it meant that not all informant questionnaires would have been completed at the same point in time as the patient ones. A review of my data from Stage One of the study showed that of the 112 people recruited, only 55 brought along an informant – a figure of just below 50% of those signed up. With hindsight, I therefore still feel this was the right decision since a

final sample of half what I did obtain would render the results meaningless. Additionally, I was already aware from previous studies (Morrow et al., 2010) that scores on the MSNQ-I were sufficiently stable and reliable over time that a difference in administration time between the informant and patient questionnaires would not damage the integrity of the data obtained. However, there was a learning point in this experience – that it would have been beneficial to explore aspects of the recruitment for the project in more explicit detail with staff at the MS Clinic prior to commencing. Given greater time resources, a pilot study would have been an ideal way to clarify such matters.

Stage two data collection began in January 2012 and presented unique challenges and learning for me. The need to continue Stage One recruitment alongside Stage Two testing meant that at times, there was a feeling of too many things going on at once and, like a circus plate spinner, having to spread myself around in order to keep all the aspects of the project running in parallel. The fact that this was achieved adequately reassures me that my organisational skills and ability to keep calm under pressure are going to be adequate when I am qualified. I also feel this aspect of the study was one of the most personally rewarding aspects of my experience as a trainee psychologist, as I was able to gain in depth experience of administering a large number of psychometric tests I had not encountered before. I found administering the DKEFS-Sorting test to be very challenging. It requires the examiner to focus intensely on the answers given and there is a great amount of recording of verbatim responses. However, as with all of the psychometrics, with practice, I feel I have become very proficient in administration. On reflection I am also pleased that my methodology explicitly allowed the PASAT to be the final test administered. I was initially cautious about its inclusion in the battery as it has a reputation for being unpleasant for the

participant but, by administering it last, I was able to have a subjective assessment of how the person I was testing had been finding the experience. This meant that, for the few people who did find the PASAT aversive, I was already prepared in what I would say to them. When I perceived that participants were fatigued I usually created an impromptu break for coffee prior to the administration of the COWAT and PASAT. I also specifically developed a set of things I would say to all participants, after they completed the PASAT, which acknowledged its difficulty and allowed those who had struggled with it to be able to leave the testing room not feeling embarrassed or ashamed. One of the aspects of my learning from the current study that will be of great value to me in my later practice will be the ability to know how to handle psychometric testing with those who are finding it difficult and aware they are not performing well.

As Stage Two of data collection continued, it became clear that it would be a struggle to obtain the full sample that was required for power. Despite booking in three participants for each day of psychometrics, there would always be at least one person who was unable to attend. By March 2012, it became clear that the full sample would not be able to be obtained and the option to use a further research assistant to speed up data collection at this stage was no longer feasible given this would require going back to ethics and R&D committees for approval and which would be unlikely to be gained before the project ended in April 2012. It seemed a point had already been reached beyond which the only option was to continue and obtain the best sample possible. This was hugely disappointing to me – not least since I had hoped to be able to publish the study and publication would now be unlikely with an underpowered sample. Hindsight is (of course) always twenty-twenty but my earlier speed in recruitment to Stage One had left me feeling that my target was achievable, but it clearly was not. However, should I wish to carry out further studies of this sort, I now have knowledge of what the

time and resource implications involved would be. I would not have obtained this insight without carrying out the project itself.

Data analysis and Writing up (Jigsaw Falling Into Place)

Having collected as much data as I could by the end of April, I made the joint decision with my supervisor that I would stop at the point I had reached and begin the process of analysing what I had and writing up the thesis. My confidence with numbers and statistics is quite high but I needed to remind myself of what I was to do. Thankfully this was not too difficult, as in the process of planning my study I had contacted Margaret Pepe – an esteemed statistician in the United States whose research studies were relevant to what I wished to achieve. Her research student had emailed me back with very straightforward instructions on how to proceed with the analysis and explaining in non-technical language what the process of combining two screening results would require me to do. Reminding myself of this email, re-acquainting myself with the statistical processes by reading texts and contacting the department's statistics adviser, was enough for me to be able to confidently proceed and complete this task relatively easily.

Writing up the thesis would have been a daunting task if seen as one task all of its own but in approaching this final and most important aspect of the process, my strategy was to break the task down into each of its' component parts and work on these one by one. In this sense, this was a task similar to completing a jigsaw – by working on it piece by piece, the task is always achievable given enough time and the big picture emerges as you go along. While extremely time intensive, the process of writing up the project has certainly been far less stressful than actually gathering the data, which was in turn, also far less stressful than obtaining ethical clearance for the project.

Conclusions

Completion of this project has been a significant learning experience for me. There are perhaps three significant areas where I have gained personally and academically from the process. First of these is the realisation of how complex the ethical approval process can be and that ensuring ethical approval is granted as quickly as possible is in itself a demanding job requiring much chasing up and checking as is reasonable. Negotiation and maintaining open communication are aspects of my skill set that have undoubtedly been enhanced through my interactions with ethical committees and research and development departments.

Secondly, the importance of adequate planning and information collection prior to commencing research is something I will most certainly carry with me to future research projects. Looking back over the process, I feel that in taking on a project that had already been approved, it was far too easy for me to assume the project was more completely defined than it was in practice. This was especially the case in respect of use of facilities at the neuropsychology service where second stage testing was carried out. I feel more discussion and negotiation should have occurred from me to them prior to starting data collection. While the department were very accommodating, I feel they should have been informed by me at an earlier stage as to how things would work, especially with regard to reimbursing them for costs of scoring sheets that I used.

The final significant aspect of my learning from the project has been the importance of co-operative working. The project simply would not have been possible without the assistance of a large number of people ranging from administrative staff in the ethical process to practitioners who allowed me to use their clinics and buildings for data collection. Without assistance from these people, the project would not have happened no matter how much I may have wanted it to. There is also the question as to

whether I should have used a research assistant for data collection. Looking back from the end-point of the project, this would almost certainly have meant that power could have been achieved but given the early success in recruiting at MS clinics, it did not seem needed. However, if conducting a study such as this in future, I would definitely involve another researcher as this would make the process less stressful and difficult for me individually.

In conclusion, the project has been an invaluable experience that has stretched and challenged me in multiple ways as well as giving me learning experiences I will carry forward into my practice as a psychologist in the future as well as for any future research I may carry out.

Appendices

Appendix A:

Search Methodology

Database	PsychINFO/ EXTRA/ ARTICLES	PsychINFO/ EXTRA/ ARTICLES	PubMed	Scopus	Web of Knowledge
Search Areas	All Fields (Peer Reviewed)	All fields (peer reviewed)	Title/Abstract		Topic
Search Terms	"Multiple Sclerosis" and "Cognitive Screening"	"Multiple Sclerosis" and "Cognitive Dysfunction" and "Screening	"Multiple Sclerosis" and "Cognitive Dysfunction" and "Screening"	"Multiple Sclerosis" (Title/Abstract /Keyword) and "Cognitive Screening" (All Fields)	"Multiple Sclerosis" and "Cognitive Dysfunction" and "Screening"
Number of artilcles produced from searches	10	10	20	30	42
Total number of articles from all searches				112	
Total number of articles after removal of duplication				70	
Total number of articles after reading abstracts to determine relevance				36	
Total number of articles after full reading with reference to specific criteria				13	

Appendix B:

Criteria for Excluding Articles

Screening of papers by reading abstracts

Criteria:

1. Paper must be available in English
2. Paper must be specifically about MS and not about other syndromes or conditions experienced prior to MS diagnosis
3. Paper must mention cognitive screening tests
4. Must be a published paper and not a letter or correction

Screening of papers by reading full text

Criteria:

1. Paper must be specifically about screening for cognitive impairment
2. Paper must explicitly test for impairment or screen for impairment using neuropsychological methods.
3. Comparison to a set of gold standard tests must be carried out.
4. Should be original research and not just a summary or review of previously published work.

Appendix C:

Details of Inclusion and Exclusion of Articles

Details of Article	Included or Excluded?	Reason for Decision
Akbar, Honarmand, Kou, Levine, Rector and Feinstein (2010)	Included	Assesses the use of MSNQ over the internet to screen for cognitive impairment. Neuropsychological tests used as gold standard
Akbar, Lobaugh and O'Connor (2010)	Excluded	Concerns MRI scan differences in cognitively impaired MS patients and not neuropsychological testing
Basso, Beason-Hazen, Lynn, Rammohan and Bornstein (1996)	Included	Concerns validation of a screening battery
Benedict (2005)	Excluded	Does not explicitly test for impairment in MS patients, this is simply an article reviewing and highlighting the use of the MSNQ
Benedict and Zivadinov (2011)	Excluded	A review of how to manage cognitive dysfunction in MS clinics
Beatty, Paul, Wilbanks, Hames, Blanco and Goodkin (1995)	Included	Concerns the use of a brief screening battery to determine impairment in comparison to a gold standard battery of neuropsychological tests
Belman-Strobl, Wuerfel, Aktas, Dorr, Wernecke, Zipp and Paul (2009)	Excluded	Compares results of screening test with Gd-enhancement MRI results which are a surrogate for the inflammatory activity in MS
Benedict, Duquin, Jurgensen, Rudick, Feitcher, Munschauer, Panzara and Weinstock-Guttman (2008)	Excluded	No comparison to a gold standard measure even though study concerns test-retest reliability of both the SDMT and MSNQ

Benedict, Cox,Thompson, Foley, Weinstock-Guttman and Munscauer (2004)	**Included**	**Replication of earlier study with larger sample. Compares screening test score with neuropsychological**
Benedict and Zivadinov (2007)	**Excluded**	**Only a review of previous work and studies with MSNQ**
Chamelain (2005)	**Excluded**	**Does not concern NP testing but using MRI to screen for impairments**
Claesson, Ytterberg, Johansson, Olmkvist and Van Koch (2007)	**Included**	**Uses a short screening test, the Free Recall and Recognition test, and checks correlation of results with other NP tests (PASAT, SDMT, MMSE)**
Dent and Lincoln (2000)	**Included**	**Tests the Brief Repeatable Battery as a screening test compared to other NP test results**
Engel, Greim and Zettel (2007)	**Excluded**	**A review of previous research studies**
Franklin, Heaton and Nelson (1988)	**Excluded**	**Concerns the correlation of MRI findings and neuropsychological tests and does not use neuropsychological tests as gold standard**
Grabner, Popotnig, Ropele, Neuper, Gorani, Petrovic, Ebner, Strasser-Fuchs, Fazekas and Enzinger (2007)	**Excluded**	**Does concern the Faces Symbol Test (a screening test) but examines brain activation patterns in MRI scans produced when patients complete it**
Julian (2011)	**Excluded**	**Only a review of cognitive impairments found in MS and not concerned with validating screening tests**
Lechner-Scott, Kerr, Spencer, Agland, Lydon and Schofield (2010)	**Included**	**Examines a screening test (ARCS) against a gold standard battery of tests**

Morrow, O'Connor, Polman, Goodman, Kappos, Lublin, Rudick, Jurgensen, Paes, Forrestal and Benedict (2010)	**Excluded**	**Uses SDMT and MSNQ as measures of impairment in a test of a drug designed to reduce impairment. Shows reliability of these tests over time but no comparison to a gold standard is carried out**
Negreiros, Mattos, Landeira-Fernandez, Paes and Alvarenga (2008)	**Included**	**Examines use of a brief screening battery to discriminate impaired RRMS patients from controls**
O'Brien, Gaudino-Goering, Shawaryn, Komaroff, Moore and DeLuca (2007)	**Included**	**Assesses the MSNQ against a range of outcome measures including a neuropsychological battery**
Olazaran, Cruz, Benito-Leon, Morales, Duque and Rivera-Navarro (2009)	**Excluded**	**Does examine use of neuropsychological tests as screening tools but only does this to derive prevalence results and does not compare screening test to a gold standard as Scherer (2007) suggests should be done**
Parmenter, Weinstock-Guttman, Garg, Munschauer and Benedict (2007)	**Included**	**Validation of SDMT as a screening tool**
Peyser (1984)	**Excluded**	**Concerns what should go into a NP assessment from writers own experience but does not validate or test a NP screening tool**
Possa (2010)	**Excluded**	**A review of the literature, not a test of neuropsychological tools**
Proseigel and Michael (1993)	**Excluded**	**A review of the literature and recommendations for rehabilitation of those with impairments**

Rao (1984)	**Excluded**	**A review article concerning how memory is affected in MS**
Rao (1995)	**Excluded**	**Only a review of past studies and not an original piece of research**
Sartori, Belliard, Chevrier, Trebon, Chaperon and Edan (2006)	**Excluded**	**This article is in French and only the abstract is translated into English**
Sartori and Edan (2006)	**Excluded**	**Only a review of the literature with recommendations of tests that might be used but no attempt to validate them in the study itself**
Scherer (2007)	**Excluded**	**A review of the literature and methods used in validating screening tests**
Scherer, Penner, Rohr et al. (2007)	**Included**	**Validation of the Faces Symbol test as screening tool in MS**
Stenager, Knudsen and Jensen (1992)	**Excluded**	**Only a one page letter summarising research and not a full report**
Vanotti, Benedict, Acion, Caceres and VANEM Workgroup (2009)	**Included**	**Validation of the MSNQ in an Argentinean population**
Wilken, Kane, Sullivan, Wallin, Usiskin, Quig, Simsarian, Saunders, Crayton, Mandler, Kerr, Reeves, Fuchs, Manning and Keller (2003)	**Included**	**Validation of the ANAM computerised screening tool in MS**
Wishart (2006)	**Excluded**	**Not an original piece of research but a review of the literature**

Appendix D:

Data Extraction Form

Data Extraction Form

Study (Reference):

Screening tool:

Methodology:

Analysis:

Sample Size:

PPMS/SPMS/RRMS/Other:

Ethnicity:

Gender

Sensitivity:

Specificity:

Accuracy (AUROC):

Other notes:

Appendix E: Quality Assessment Tool

Please refer to www.phru.nhs.uk/casp/casp.htm

Appendix F:

Statement of Epistemological Position

Epistemological Position

The study was conducted taking the positivist approach to knowledge in order to test hypotheses and either accept or reject them. Statistical methods were used to this end in order to establish whether or not the results could have occurred by chance. Additionally the use of neuropsychological tests implicitly requires the assumption that psychological concepts such as memory, attention and executive function relate to genuine measurable entities and abilities possessed by human brain.

Appendix G:

Participant

Information Sheet

INFORMATION FOR PARTICIPANTS

CAN YOU HELP?

SCREENING FOR COGNITIVE PROBLEMS IN PEOPLE WITH MULTIPLE SCLEROSIS

INVESTIGATORS: MR. R. BURTON, PROF. M. WANG

Researchers at Leicestershire NHS Trusts and the University of Leicester are conducting a study looking into improving services for people with Multiple Sclerosis in the region. We are looking for people who would be willing to participate in the study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take your time to read the following information carefully and discuss it with relatives, friends or your GP if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the research about?

We would like to see whether using a combination of a brief questionnaire and a 5-minute test can accurately identify those people who might benefit most from a referral to, and support from our specialist Neuropsychology Service here at the Leicester General Hospital. The brief questionnaire comes in two forms – one for the patient and the other for the patient's carer or next of kin.

Why have I been chosen?

We are asking everyone who attends the Outpatient Multiple Sclerosis Clinics to take part in this study.

Do I have to take part?

It is up to you whether or not to take part. If you decide to take part you will be given this information sheet to keep and be asked to sign a consent form. If you decide to take part you are still free to withdraw at any time and without giving a reason. This will not affect the standard of care you receive.

What will I have to do?

There are two stages to this study.

Stage One: This would take place at your next appointment at the Outpatient Multiple Sclerosis Clinic. We will ask you to complete a short tick box questionnaire to help us examine your current levels of psychological and intellectual functioning. In addition we will ask you to complete a simple test of your speed of thinking, which will take no more than five minutes. If it is possible we would ask your carer or next of kin to complete a version of the questionnaire as well.
Stage Two: If you agree to take part in the second stage of this study, you will be asked to come back to the Leicester General Hospital within the next few weeks.

You will then meet with one of our colleagues from the Neuropsychology Service who will take you through a further series of brief assessments. This stage will take up to a maximum of **two hours,** with regular rest breaks. To be fair we are required to select people at random, which means if you choose to volunteer for stage two, you will only have a 1 in 3 chance of being chosen to participate.

What are the possible disadvantages and risks of taking part?
There are no particular risks involved in taking part in the study. Those volunteers who are randomly selected for stage two of the study need to come back to the Leicester General Hospital a week or two after their Outpatient appointment.

What are the possible benefits of taking part?
With the information gained from this study we hope to develop a better understanding of the psychological and intellectual problems affecting people with Multiple Sclerosis. These difficulties often remain unrecognised but can cause people a lot of stress in their everyday lives. The information we get from this study may help us to treat and support patients with Multiple Sclerosis more effectively.
Patients who volunteer for stage two of the study will be given a brief report of their assessment results, at their request. We cannot guarantee there will be personal benefit to you from taking part in the study.

What if something goes wrong?
If you are unhappy or wish to complain about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanism will be available to you. You can raise your complaint personally with the researcher or your clinician or contact the Patient Information Liaison Service (PILS):

Freephone: 0808 178 8337
Email: pils.complaints.compliments@uhl-tr.nhs.uk
Address: Trust Headquarters, Gwendolen House, Gwendolen Road, Leicester LE5 4QF

Alternatively, you can make a complaint through an independent body such as the Independent Complaints Advocacy Service (ICAS) whose East Midlands service can be contacted on 0300 456 8347 or your local citizens advice bureau.

Will my taking part in this study be kept confidential?
If you consent to take part your medical records may be looked at by the researchers for the purpose of gathering data and analysing results. All information that is collected about you during the course of the research will be kept strictly confidential; any information about you that leaves the hospital will have your personal details removed so that you cannot be identified from it.

What will happen to the results of this research study?
The study will contribute towards the principal investigator's Doctorate in Clinical Psychology qualification. We plan to publish the results of this study in a scientific journal. However, you will not be identified in any report/ publication. If you would like to receive a summary of the published results, please let us know at your Outpatient appointment.

Who has reviewed the Study?
The Leicestershire, Northamptonshire and Rutland Research Ethics Committee 2 has reviewed the study.

Further Information
If you have any further questions about this study you may contact: Richard Burton, at Clinical Neuropsychology Services, Leicester General Hospital, LE5 4PW Tel: 0116 258 4956.

Thank you for taking time to read this. If you agree to take part, please bring the enclosed consent form with you to your next appointment.

Appendix H:

Participant Consent

Form

Version 6: 22.07.2011

University Hospitals of Leicester NHS
NHS Trust

RECRUITING AT LEICESTER GENERAL HOSPITAL, LEICESTER LE5 4PW

CONSENT FORM

SCREENING FOR COGNITIVE PROBLEMS IN PEOPLE WITH MULTIPLE SCLEROSIS

INVESTIGATORS: MR. R. BURTON, PROF. M. WANG

The patient should complete the whole of this sheet himself/herself.

Please initial in the boxes as applies:

I have read & understood the patient information sheet (version 5: 22.07.2011) ☐

I have had sufficient opportunity to ask questions & discuss the study ☐

All my questions been answered satisfactorily ☐

I understand my carer or next of kin will complete a brief tick box questionnaire about my current psychological and intellectual functioning ☐

I understand that I am free to withdraw from the study at any time, without having to give a reason and that my decision whether or not to participate will not affect my future medical care. ☐

The named researchers can view my medical records for the purpose of this study ☐

I would like to receive information regarding the results of this study ☐

I understand that relevant sections of my medical notes and data collected during the study may be looked at by individuals from the University of Leicester, from regulatory authorities, from Leicestershire Partnership Trust or from University Hospitals of Leicester, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data and records. ☐

Who have you spoken to Dr/Mrs/Ms/Mr …….....

I agree to take part in this study

Signature (patient) Date:

Name (In block capitals)

Signature (carer – if appropriate) Date:

Name (In block capitals)

I have explained the study to the above patient and he/she has indicated his/her willingness to take part

Signature: Date:
Name (In block capitals):

Appendix I:

Informant

Information Sheet

INFORMATION FOR NEXT OF KIN/ RELATIVES/ CARERS

CAN YOU HELP?

SCREENING FOR COGNITIVE PROBLEMS IN PEOPLE WITH MULTIPLE SCLEROSIS

INVESTIGATORS: MR. R. BURTON, PROF. M. WANG

Researchers at Leicestershire NHS Trusts and the University of Leicester are conducting a study looking into improving services for people with Multiple Sclerosis in the region. We are looking for people who would be willing to participate in this study. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take your time to read the following information carefully and discuss it with relatives, friends or your relative's GP if you wish. Ask us if there is anything that is not clear or if you would like more information.

What is the research about?
We would like to see whether using a combination of a brief questionnaire and a 5-minute test can accurately identify those people who might benefit most from a referral to, and support from our specialist Neuropsychology Service here at the Leicester General Hospital. The brief questionnaire comes in two forms – one for the patient and the other for the patient's carer or next of kin.

Why has my relative been chosen?
We are asking everyone who attends the Outpatient Multiple Sclerosis Clinics to take part in this study.

Does my relative have to take part?
It is entirely up to each individual whether or not to take part.

What will my relative have to do?
There are two stages to this study. You can choose for you and your relative just to take part in stage one or both stages of the study.

Stage One: This involves you completing a short tick box questionnaire to help us examine what you consider to be your relative's, or the person you care for's, current levels of psychological and intellectual functioning.

Stage Two: We would not require your direct participation in the second part of the study. If the person you are caring for, or your relative, agrees to take part in stage two of the study they would be asked to return to Leicester General Hospital within the next few weeks and you would be welcome to attend with them.

They will then meet with one of our colleagues from the Neuropsychology Service who will take them through a further series of brief assessments. This stage will take a maximum of two hours, with regular rest breaks. To be fair we are required to select people at random, which means there will be have a 1 in 3 chance of your relative or the person you care for being chosen to participate.

What are the possible disadvantages and risks of taking part?
There are no particular risks involved in taking part in the study. Those volunteers who are randomly selected for stage two of the study will be asked to come back to the Leicester General Hospital a week or two after their Outpatient appointment.

What are the possible benefits of taking part?
With the information gained from this study we hope to develop a better understanding of the psychological and intellectual problems affecting people with Multiple Sclerosis. These difficulties often remain unrecognised but can cause people a lot of stress in their everyday lives. Information we get from this study may help us to treat and support patients with Multiple Sclerosis more effectively. We cannot guarantee there will be personal benefit to people taking part in the study.

What if something goes wrong?
If you are unhappy or wish to complain about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanism will be available to you. You can raise your complaint personally with the researcher or your clinician or contact the Patient Information Liaison Service (PILS):

Freephone: 0808 178 8337
Email: pils.complaints.compliments@uhl-tr.nhs.uk

Address: Trust Headquarters, Gwendolen House, Gwendolen Road, Leicester LE5 4QF

Alternatively, you can make a complaint through an independent body such as the Independent Complaints Advocacy Service (ICAS) whose East Midlands service can be contacted on 0300 456 8347 or your local citizens advice bureau.

Will my taking part in this study be kept confidential?
All information that is collected during the course of the research will be kept strictly confidential. Any information that leaves the hospital will have your personal details removed so that you cannot be identified from it.

What will happen to the results of this research study?
The study will contribute towards the principal investigator's Doctorate in Clinical Psychology qualification. We plan to publish the results of this study in a scientific journal. However, you will not be identified in any report/ publication. If you would like to receive a summary of the published results, please let us know at your Outpatient appointment.

Who has reviewed the Study?
The Leicestershire, Northamptonshire and Rutland Research Ethics Committee 2 has reviewed the study.

Further Information
If you have any further questions about this study you may contact: Richard Burton at Clinical Neuropsychology Services, Leicester General Hospital, LE5 4PW. Tel: 0116 258 4956.

Thank you for taking time to read this. If you agree to take part, please bring the enclosed consent form (for realtives/next of kin/ carers) with you to the MS Outpatients Clinic.

Appendix J:

Informant Consent

Form

Version 6: 22.07.2011

University Hospitals of Leicester NHS
NHS Trust

RECRUITING AT LEICESTER GENERAL HOSPITAL, LEICESTER LE5 4PW

CONSENT FORM (FOR NEXT OF KIN/RELATIVES/CARERS)

SCREENING FOR COGNITIVE PROBLEMS IN PEOPLE WITH MULTIPLE SCLEROSIS

INVESTIGATORS: MR. R. BURTON, PROF. M. WANG

You should complete the whole of this sheet yourself.

Please initial in the boxes as applies:

I have read & understood the information sheet for next of kin/relatives/carers (version 5: 22.07.2011) ☐

I have had sufficient opportunity to ask questions & discuss the study ☐

All my questions been answered satisfactorily ☐

I understand that I am free to withdraw from the study at any time, without having to give a reason and that my decision whether or not to participate will not affect mine or anyone else's medical care. ☐

I would like to receive information regarding the results of this study ☐

I understand that data collected during the study may be looked at by individuals from the University of Leicester, from regulatory authorities, from Leicestershire Partnership Trust or from University Hospitals of Leicester, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my data ☐

Who have you spoken to Dr/Mrs/Ms/Mr …….....

I agree to take part in this study

Signature (participant): Date:

Name (In block capitals):

I have explained the study to the above person and he/she has indicated his/her willingness to take part

Signature: Date:

Name (In block capitals):

Appendix K:

Multiple Sclerosis Neuropsychological Questionnaire

(Removed due to professional code of conduct)

Appendix L:

Symbol Digit Modalities Test

(Removed due to professional code of conduct)

Appendix M:

Judgement of Line Orientation Test (JLO)

(Removed due to professional code of conduct)

Appendix N:

Controlled Oral Word Association Test (COWAT)

(Removed due to professional code of conduct)

Appendix O:

California Verbal

Learning

Test – II (CVLT-II)

(Removed due to professional code of conduct)

Appendix P:

Brief Visuospatial Memory Test – Revised (BVMT-R)

(Removed due to professional code of conduct)

Appendix Q:

Delis-Kaplan Executive Function System – Sorting Test

(Removed due to professional code of conduct)

Appendix R:

Paced Auditory Serial Addition Test (PASAT)

(Removed due to professional code of conduct)

Appendix S:

Beck Depression Inventory – Fast Screen

(Removed due to professional code of conduct)

Appendix T:

Fatigue Impact Scale

(Removed due to professional code of conduct)

Appendix U:

Letters Regarding

Ethical Approval

Process

(Removed due to confidentiality requirements)

Appendix V:

Chronology of Research Process

Chronology of Research Process

Date	Aspect of process
December 2009	Research fair attended – information provided about potential project
	Meet with Joanne Kirkland to discuss project
January – February 2010	Carry out first year literature review in research area
March 2010	Final decision made on research area
June 2010	First draft of research proposal reviewed internally by DClinPsy course staff
July 2010	Meeting with MS Specialist nurses at Neurology Outpatients department to confirm study will be feasible
August – November 2010	Revisions to research proposal made
December 2010 – January 2011	Completion of IRAS form online in preparation for submission to ethics
January 2011	Submission of ethics application to Research Ethics Committee
February 2011	Research Ethics Committee (REC) panel meeting
	Feedback from REC panel received
April 2011	Response to REC sent
	Positive ethical opinion received from REC
	Application submitted to R&D department
May 2011	Major amendment to protocol submitted to REC on request of

	R&D
June 2011	Positive ethical opinion received from REC for major amendment to protocol
June – August 2011	Two further minor amendments made to information forms and letters to be used in study
	Provision of assorted documents to R&D department on request
August 2011	Permission to commence study received from R&D department
September 2011	First information forms sent to potential participants ahead of their attendance at MS clinic
October 2011 – April 2012	Data collection at MS clinic, neuropsychology service
	Data entry into SPSS
December 2011	Writing of first draft of Literature Review
January 2011	Writing of first draft of Methodology for research report
March 2011	Writing of second draft of Methodology for research report
April 2011	Writing of second draft of Literature Review
	Writing of first draft of Introduction for research report
April – May 2011	Data analysis
May 2011	Completion of Critical Appraisal first draft
	Writing of remaining aspects of research report
	Submission of soft-bound thesis to University of Leicester

Appendix W:

Guidelines to Authors for Journal of Neurology

Please refer to www.springer.com

Appendix X:

SPSS Output for Comparison of Impaired and Non-Impaired Groups

Chi-Squared Tests – Gender of Participants

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Gender of Participant * MACFIMS Outcome - Impaired / Not Impaired	26	100.0%	0	.0%	26	100.0%

Gender of Participant * MACFIMS Outcome - Impaired / Not Impaired Crosstabulation

Count

		MACFIMS Outcome - Impaired / Not Impaired		
		Not Impaired	Impaired	Total
Gender of Participant	Male	7	4	11
	Female	8	7	15
Total		15	11	26

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.276[a]	1	.599		
Continuity Correction[b]	.015	1	.902		
Likelihood Ratio	.278	1	.598		
Fisher's Exact Test				.701	.452
N of Valid Cases	26				

a. 1 cells (25.0%) have expected count less than 5. The minimum expected count is 4.65.

b. Computed only for a 2x2 table

Chi-Squared Tests – Ethnicity of Participants

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Ethnicity of Participant * MACFIMS Outcome - Impaired / Not Impaired	26	100.0%	0	.0%	26	100.0%

Ethnicity of Participant * MACFIMS Outcome - Impaired / Not Impaired Crosstabulation

Count

		MACFIMS Outcome - Impaired / Not Impaired		
		Not Impaired	Impaired	Total
Ethnicity of Participant	White British	12	9	21
	Mixed	0	1	1
	White Other	1	0	1
	Asian British	2	1	3
Total		15	11	26

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.199[a]	3	.532
Likelihood Ratio	2.925	3	.403
N of Valid Cases	26		

a. 6 cells (75.0%) have expected count less than 5. The minimum expected count is .42.

Chi-Squared Tests – Diagnosis of Participants

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Type of MS Diagnosis * MACFIMS Outcome - Impaired / Not Impaired	26	100.0%	0	.0%	26	100.0%

Type of MS Diagnosis * MACFIMS Outcome - Impaired / Not Impaired Crosstabulation

Count

		MACFIMS Outcome - Impaired / Not Impaired		
		Not Impaired	Impaired	Total
Type of MS Diagnosis	Relapsing-Remitting	6	9	15
	Secondary Progressive	5	0	5
	Primary Progressive	4	2	6
Total		15	11	26

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.788[a]	2	.055
Likelihood Ratio	7.597	2	.022
N of Valid Cases	26		

a. 4 cells (66.7%) have expected count less than 5. The minimum expected count is 2.12.

T-tests – Age of Participant, Years of Education, Years since diagnosis

Group Statistics

	MACFIMS Outcome - Impaired / Not Impaired	N	Mean	Std. Deviation	Std. Error Mean
Age of Participant	Not Impaired	15	50.2553	9.80162	2.53077
	Impaired	11	48.4682	12.69178	3.82671
Number of Years of Education	Not Impaired	15	14.800	3.7023	.9559
	Impaired	11	12.273	2.1950	.6618
Number of Years since MS diagnosis	Not Impaired	15	6.167	3.5740	.9228
	Impaired	11	11.864	11.5868	3.4936

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Age of Participant	Equal variances assumed	.128	.724	.406	24	.689	1.78715	4.40532	-7.30499	10.87929
	Equal variances not assumed			.390	18.177	.701	1.78715	4.58787	-7.84489	11.41919
Number of Years of Education	Equal variances assumed	8.776	.007	2.013	24	.055	2.5273	1.2555	-.0640	5.1185
	Equal variances not assumed			2.174	23.181	.040	2.5273	1.1627	.1231	4.9314
Number of Years since MS diagnosis	Equal variances assumed	4.708	.040	-1.803	24	.084	-5.6970	3.1605	-12.2200	.8260
	Equal variances not assumed			-1.577	11.405	.142	-5.6970	3.6134	-13.6157	2.2218

Appendix Y:

SPSS Output for ROC Analysis of Individual Tests

ROC Analysis – MSNQ-P

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	11
Negative	15

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Score on MSNQ-Patient

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.785	.098	.015	.594	.976

The test result variable(s): Score on MSNQ-Patient has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Score on MSNQ-Patient

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
7.00	1.000	1.000
10.00	1.000	.933
12.50	1.000	.867
13.50	1.000	.800
15.00	1.000	.733
17.00	.818	.667
18.50	.818	.600
19.50	.727	.600
20.50	.727	.533
22.00	.727	.467
24.00	.727	.400
26.00	.727	.333
28.00	.727	.200
30.00	.727	.133
31.50	.636	.133
32.50	.545	.067
33.50	.455	.067
36.50	.364	.067
40.00	.364	.000
42.50	.273	.000
44.50	.182	.000
46.00	.091	.000
48.00	.000	.000

The test result variable(s): Score on MSNQ-Patient has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

ROC Analysis – MSNQ-I

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	9
Negative	11
Missing	6

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Score on MSNQ-Informant

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.652	.126	.254	.405	.898

The test result variable(s): Score on MSNQ-Informant has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Score on MSNQ-Informant

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
7.00	1.000	1.000
9.00	1.000	.909
11.00	.889	.818
13.00	.889	.727
14.50	.778	.545
16.50	.667	.455
20.00	.556	.364
22.50	.444	.364
24.00	.444	.273
26.00	.333	.273
30.50	.333	.182
36.00	.333	.091
39.00	.222	.091
41.00	.222	.000
43.50	.111	.000
46.00	.000	.000

The test result variable(s): Score on MSNQ-Informant has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

ROC Analysis – SDMT

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	11
Negative	15

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Inverted score on SDMT

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.712	.103	.069	.510	.914

The test result variable(s): Inverted score on SDMT has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Inverted score on SDMT

Positive if Greater Than or Equal To[a]	Sensitivity	1 – Specificity
40.00	1.000	1.000
44.50	1.000	.933
50.50	1.000	.800
54.00	1.000	.733
55.50	.909	.667
57.50	.818	.667
59.50	.818	.600
60.50	.727	.467
63.00	.727	.333
66.00	.545	.333
67.50	.545	.267
68.50	.545	.200
70.00	.455	.200
71.50	.364	.067
72.50	.273	.067
74.00	.182	.067
81.50	.091	.067
88.50	.000	.067
90.00	.000	.000

The test result variable(s): Inverted score on SDMT has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Appendix Z:

SPSS Output for

Logistic Regressions

Logistic Regression – MSNQ-P and SDMT

Without Interaction of Variables

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	26	100.0
	Missing Cases	0	.0
	Total	26	100.0
Unselected Cases		0	.0
Total		26	100.0

a. If weight is in effect, see classification table for the total number of cases.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.888	2	.019
	Block	7.888	2	.019
	Model	7.888	2	.019

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	msnqp	.102	.051	3.961	1	.047	1.107
	sdmt	-.041	.045	.824	1	.364	.960
	Constant	-1.137	2.691	.179	1	.673	.321

a. Variable(s) entered on step 1: msnqp, sdmt.

Including Interaction of Variables

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	26	100.0
	Missing Cases	0	.0
	Total	26	100.0
Unselected Cases		0	.0
Total		26	100.0

a. If weight is in effect, see classification table for the total number of cases.

Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	9.292	3	.026
	Block	9.292	3	.026
	Model	9.292	3	.026

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	msnqp	.448	.342	1.720	1	.190	1.565
	sdmt	.144	.170	.718	1	.397	1.155
	msnqp by sdmt	-.008	.007	1.152	1	.283	.992
	Constant	-9.453	8.163	1.341	1	.247	.000

a. Variable(s) entered on step 1: msnqp, sdmt, msnqp * sdmt .

Logistic Regression - MSNQ-I & SDMT

Without Interaction of Variables

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	20	76.9
	Missing Cases	6	23.1
	Total	26	100.0
Unselected Cases		0	.0
Total		26	100.0

a. If weight is in effect, see classification table for the total number of cases.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3.892	2	.143
	Block	3.892	2	.143
	Model	3.892	2	.143

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	msnqi	.027	.045	.362	1	.547	1.028
	sdmt	-.086	.061	1.986	1	.159	.917
	Constant	3.202	3.291	.947	1	.331	24.592

a. Variable(s) entered on step 1: msnqi, sdmt.

Including Interaction of Variables

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	20	76.9
	Missing Cases	6	23.1
	Total	26	100.0
Unselected Cases		0	.0
Total		26	100.0

a. If weight is in effect, see classification table for the total number of cases.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	7.713	3	.052
	Block	7.713	3	.052
	Model	7.713	3	.052

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	msnqi	.828	.590	1.969	1	.161	2.288
	sdmt	.234	.216	1.168	1	.280	1.263
	msnqi by sdmt	-.018	.013	2.018	1	.155	.982
	Constant	-11.408	10.391	1.205	1	.272	.000

a. Variable(s) entered on step 1: msnqi, sdmt, msnqi * sdmt .

Appendix AA:

SPSS Output for

ROC Analysis on

Combined Scores

ROC Analysis – MSNQ-P & SDMT Combined Scores

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	11
Negative	15

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Risk Score (Combined) from SDMT & MSNQ-P

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.773	.097	.020	.583	.963

The test result variable(s): Risk Score (Combined) from SDMT & MSNQ-P has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Risk Score (Combined) from SDMT & MSNQ-P

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-3.3710	1.000	1.000
-2.3620	1.000	.933
-2.0565	1.000	.867
-1.6990	.909	.800
-1.5750	.909	.733
-1.4920	.909	.667
-1.4110	.909	.600
-1.2810	.818	.600
-1.1705	.818	.533
-1.0045	.818	.467
-.8605	.727	.467
-.7685	.727	.400
-.5440	.727	.333
-.3310	.727	.267
-.1275	.727	.200
.0460	.636	.200
.1105	.545	.200
.1620	.545	.133
.2925	.455	.133
.5685	.455	.067
.7820	.364	.067
1.2210	.364	.000
1.7730	.273	.000
2.0175	.182	.000
2.2740	.091	.000
3.4490	.000	.000

The test result variable(s): Risk Score (Combined) from SDMT & MSNQ-P has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

ROC Analysis – MSNQ-I & SDMT Combined Scores

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	9
Negative	11
Missing	6

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Combined Score SDMT and MSNQ-I

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.808	.103	.020	.605	1.000

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Combined Score SDMT and MSNQ-I

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-4.0740	1.000	1.000
-2.8940	1.000	.909
-2.3270	1.000	.818
-1.8590	.889	.818
-1.5710	.889	.727
-1.3190	.889	.636
-1.1030	.889	.545
-.8600	.889	.455
-.7520	.889	.364
-.6620	.778	.364
-.5630	.667	.364
-.5000	.667	.273
-.4820	.667	.182
-.4550	.667	.091
-.2930	.556	.091
.1030	.444	.091
.9400	.444	.000
2.2360	.333	.000
3.5050	.222	.000
4.8730	.111	.000
6.6920	.000	.000

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Appendix AB:

SPSS Output for

Correlational

Analyses

MSNQ and BDI-FS

Correlations

			Score on MSNQ-Patient	Score on Beck Depression Inventory - Fast Screen
Spearman's rho	Score on MSNQ-Patient	Correlation Coefficient	1.000	.684**
		Sig. (1-tailed)	.	.000
		N	26	26
	Score on Beck Depression Inventory - Fast Screen	Correlation Coefficient	.684**	1.000
		Sig. (1-tailed)	.000	.
		N	26	26

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Score on MSNQ-Informant	Score on Beck Depression Inventory - Fast Screen
Spearman's rho	Score on MSNQ-Informant	Correlation Coefficient	1.000	.616**
		Sig. (1-tailed)	.	.002
		N	20	20
	Score on Beck Depression Inventory - Fast Screen	Correlation Coefficient	.616**	1.000
		Sig. (1-tailed)	.002	.
		N	20	26

**. Correlation is significant at the 0.01 level (1-tailed).

MSNQ and FIS

Correlations

			Score on MSNQ-Patient	Fatigue Impact Scale - Cognitive Score
Spearman's rho	Score on MSNQ-Patient	Correlation Coefficient	1.000	.858**
		Sig. (1-tailed)	.	.000
		N	26	26
	Fatigue Impact Scale - Cognitive Score	Correlation Coefficient	.858**	1.000
		Sig. (1-tailed)	.000	.
		N	26	26

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Score on MSNQ-Informant	Fatigue Impact Scale - Cognitive Score
Spearman's rho	Score on MSNQ-Informant	Correlation Coefficient	1.000	.642**
		Sig. (1-tailed)	.	.001
		N	20	20
	Fatigue Impact Scale - Cognitive Score	Correlation Coefficient	.642**	1.000
		Sig. (1-tailed)	.001	.
		N	20	26

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Score on MSNQ-Patient	Fatigue Impact Scale - Physical Score
Spearman's rho	Score on MSNQ-Patient	Correlation Coefficient	1.000	.774**
		Sig. (1-tailed)	.	.000
		N	26	26
	Fatigue Impact Scale - Physical Score	Correlation Coefficient	.774**	1.000
		Sig. (1-tailed)	.000	.
		N	26	26

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Score on MSNQ-Informant	Fatigue Impact Scale - Physical Score
Spearman's rho	Score on MSNQ-Informant	Correlation Coefficient	1.000	.573**
		Sig. (1-tailed)	.	.004
		N	20	20
	Fatigue Impact Scale - Physical Score	Correlation Coefficient	.573**	1.000
		Sig. (1-tailed)	.004	.
		N	20	26

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Score on MSNQ-Patient	Fatigue Impact Scale - Social Score
Spearman's rho	Score on MSNQ-Patient	Correlation Coefficient	1.000	.758**
		Sig. (1-tailed)	.	.000
		N	26	26
	Fatigue Impact Scale - Social Score	Correlation Coefficient	.758**	1.000
		Sig. (1-tailed)	.000	.
		N	26	26

**. Correlation is significant at the 0.01 level (1-tailed).

Correlations

			Score on MSNQ-Informant	Fatigue Impact Scale - Social Score
Spearman's rho	Score on MSNQ-Informant	Correlation Coefficient	1.000	.656**
		Sig. (1-tailed)	.	.001
		N	20	20
	Fatigue Impact Scale - Social Score	Correlation Coefficient	.656**	1.000
		Sig. (1-tailed)	.001	.
		N	20	26

**. Correlation is significant at the 0.01 level (1-tailed).

Appendix AC:

SPSS Output for

ROC Analyses using

BDI-FS and FIS-

Cognitive scores

ROC Analysis – BDI-FS

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	11
Negative	15

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Score on Beck Depression Inventory - Fast Screen

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.664	.111	.161	.445	.882

The test result variable(s): Score on Beck Depression Inventory - Fast Screen has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Score on Beck Depression Inventory - Fast Screen

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
-1.00	1.000	1.000
.50	.909	.867
1.50	.909	.733
2.50	.818	.733
3.50	.727	.533
4.50	.636	.333
5.50	.545	.267
6.50	.455	.200
7.50	.273	.133
9.00	.273	.067
11.00	.182	.067
13.50	.091	.000
16.00	.000	.000

The test result variable(s): Score on Beck Depression Inventory - Fast Screen has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

ROC Analysis – FIS-Cognitive

Case Processing Summary

MACFIMS Outcome - Impaired / Not Impaired	Valid N (listwise)
Positive[a]	11
Negative	15

Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.

a. The positive actual state is Impaired.

Area Under the Curve

Test Result Variable(s):Fatigue Impact Scale - Cognitive Score

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.818	.092	.006	.637	.999

The test result variable(s): Fatigue Impact Scale - Cognitive Score has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Coordinates of the Curve

Test Result Variable(s):Fatigue Impact Scale - Cognitive Score

Positive if Greater Than or Equal To[a]	Sensitivity	1 - Specificity
.00	1.000	1.000
1.50	1.000	.933
2.50	.909	.867
4.00	.909	.800
5.50	.909	.733
6.50	.909	.600
7.50	.818	.467
10.50	.818	.400
14.50	.818	.333
17.50	.818	.267
19.50	.727	.267
20.50	.727	.200
21.50	.636	.200
22.50	.636	.067
24.00	.545	.067
25.50	.455	.000
27.00	.364	.000
28.50	.273	.000
30.00	.182	.000
32.50	.091	.000
35.00	.000	.000

The test result variable(s): Fatigue Impact Scale - Cognitive Score has at least one tie between the positive actual state group and the negative actual state group.

a. The smallest cutoff value is the minimum observed test value minus 1, and the largest cutoff value is the maximum observed test value plus 1. All the other cutoff values are the averages of two consecutive ordered observed test values.

Appendix AD:

SPSS Output for Logistic Regression including BDI-FS and FIS scores

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	26	100.0
	Missing Cases	0	.0
	Total	26	100.0
Unselected Cases		0	.0
Total		26	100.0

a. If weight is in effect, see classification table for the total number of cases.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	12.173	6	.058
	Block	12.173	6	.058
	Model	12.173	6	.058

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	msnqp	.087	.086	1.029	1	.310	1.091
	sdmt	-.041	.071	.339	1	.560	.960
	bdifs	.024	.199	.015	1	.904	1.024
	fiscog	.231	.137	2.845	1	.092	1.260
	fisphys	-.024	.125	.036	1	.850	.977
	fissoc	-.089	.094	.911	1	.340	.914
	Constant	-1.318	3.691	.128	1	.721	.268

a. Variable(s) entered on step 1: msnqp, sdmt, bdifs, fiscog, fisphys, fissoc.

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	20	76.9
	Missing Cases	6	23.1
	Total	26	100.0
Unselected Cases		0	.0
Total		26	100.0

a. If weight is in effect, see classification table for the total number of cases.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	9.346	6	.155
	Block	9.346	6	.155
	Model	9.346	6	.155

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	msnqi	-.025	.066	.146	1	.702	.975
	sdmt	-.090	.101	.792	1	.373	.914
	bdifs	.139	.252	.305	1	.581	1.149
	fiscog	.152	.140	1.167	1	.280	1.164
	fisphys	.158	.179	.782	1	.376	1.171
	fissoc	-.107	.105	1.034	1	.309	.898
	Constant	1.164	5.570	.044	1	.835	3.202

a. Variable(s) entered on step 1: msnqi, sdmt, bdifs, fiscog, fisphys, fissoc.