Journal of the International Neuropsychological Society (2010), **16**, 6–16.

doi:10.1017/S1355617709990750

The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS)

BRETT A. PARMENTER,[1] S. MARC TESTA,[2] DAVID J. SCHRETLEN,[2]
BIANCA WEINSTOCK-GUTTMAN,[3] AND RALPH H. B. BENEDICT[3]

[1]Department of Psychology, Western State Hospital, Tacoma, Washington
[2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland
[3]Department of Neurology, Jacobs Neurological Institute, State University of New York at Buffalo, School of Medicine and Biomedical Sciences, Buffalo, New York

(RECEIVED March 25, 2009; FINAL REVISION July 17, 2009; ACCEPTED July 20, 2009)

Abstract

The Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS) is a consensus neuropsychological battery with established reliability and validity. One of the difficulties in implementing the MACFIMS in clinical settings is the reliance on manualized norms from disparate sources. In this study, we derived regression-based norms for the MACFIMS, using a unique data set to control for standard demographic variables (i.e., age, age^2, sex, education). Multiple sclerosis (MS) patients (n = 395) and healthy volunteers (n = 100) did not differ in age, level of education, sex, or race. Multiple regression analyses were conducted on the performance of the healthy adults, and the resulting models were used to predict MS performance on the MACFIMS battery. This regression-based approach identified higher rates of impairment than manualized norms for many of the MACFIMS measures. These findings suggest that there are advantages to developing new norms from a single sample using the regression-based approach. We conclude that the regression-based norms presented here provide a valid alternative to identifying cognitive impairment as measured by the MACFIMS. (*JINS*, 2010, *16*, 6–16.)

Keywords: Standard scores, Immunologic disease, Neuropsychology, Normalization, psychometrics, Brain

INTRODUCTION

Two types of norms are typically used to interpret performance on neuropsychological measures: discrete and continuous (Zachary & Gorsuch, 1985). Discrete norms include sets of descriptive statistics for specific age groups (Klein, Foerster, & Hartnegg, 2007), although norms can be based on rather arbitrary age bands. For example, tables may be divided by five-year intervals (e.g., norms for individuals between the ages of 20 and 24 years, 25 and 29 years, and so on), ten-year intervals, or any other age range. These norms are appropriate if the mean and standard deviation for the normative sample group approximate the mean and standard deviation for the *true* population and if the raw scores are normally distributed (Zachary & Gorsuch, 1985). One problem with discrete norms is that an individual's apparent performance can shift depending on which age band is used, even though the raw score remains the same. Zachary and Gorsuch (1985) noted that a person's IQ score on the Wechsler Adult Intelligence Scale–Revised (WAIS-R) could increase up to six points by aging a single day when their raw test scores pass from comparison with the 25–34 to the 35–44 year age group. Because of this instability, Zachary and Gorsuch recommended using continuous norms as an alternative to discrete norms.

An alternative approach is to derive continuous norms using multiple regression equations. Predictor variables can vary, but usually are specific demographic variables that have been shown to affect performance on neuropsychological measures. Such demographic variables always include age and education (Crawford & Allan, 1997; Heaton, Ryan, Grant, & Matthews, 1996; Leckliter & Matarazzo, 1989), as well as sex

Correspondence and reprint requests to: Ralph H. B. Benedict, Ph.D., Department of Neurology, 100 High Street (D-6), Buffalo, New York 14203. E-mail: benedict@buffalo.edu

(Crawford & Allan, 1997; Leckliter & Matarazzo, 1989; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2005; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006a), although sex often has little impact on cognitive performance (Heaton, Ryan, Grant, & Matthews, 1996; Sherrill-Pattison, Donders, & Thompson, 2000). Age-squared can also be added as a predictor variable to evaluate the influence of nonlinear age effects on normal cognitive performance (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006a,b).

Several investigators argue that norms based on multiple regression equations are useful insofar as they allow an individual's predicted score on a measure to reflect *specific* demographic characteristics (Heaton, Avitable, Grant, & Matthews, 1999; Crawford & Howell, 1998). Identification of the most relevant demographic variables can be a challenge (Van Breukelen & Vlaeyen, 2005), but using such equations has reduced demographic biases in the raw data derived from the Boston Naming Test (Heaton, Avitable, Grant, & Matthews, 1999), Rey Auditory Verbal Learning Test (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2005), and Stroop Color-Word Test (Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006b), to name a few. This method has also been used to minimize such demographic biases in a large cognitive battery consisting of 19 individual measures (Schretlen et al., 2007).

If the regression-based approach can control for the influence of demographic variables, might the same approach apply to control for physical or neurological disability factors? Indeed, a few authors have suggested including not only demographic variables, but also neurologic variables in "norms" for specific clinical samples (Vanderploeg et al., 1997), such as length of coma or Glasgow Coma Scale scores in interpreting traumatic brain injury data. Sherrill-Pattison, Donders, & Thompson (2000) found that the severity of brain injury, as defined by length of coma, was a significant predictor of performance on neuropsychological tests.

In the present study, we endeavored to calculate regression-based norms that would assist in the clinical evaluation of patients with multiple sclerosis (MS). Like Schretlen et al. (2007), we aimed to derive data that allow for the control (and assessment) of such confounding variables as age and education. In addition, we sought to evaluate the influence of other clinical factors such as depression, dysarthria, and upper-extremity motor function. We focused on the Minimal Assessment of Cognitive Function in Multiple Sclerosis (MACFIMS; Benedict et al., 2002), a collection of tests based on consensus opinion designed to briefly evaluate the "principle features of MS-related cognitive dysfunction" (p. 382).

Cognitive impairment affects between 43–60% of patients with MS (Benedict et al, 2006; Rao, Leo, Bernardin, & Unverzagt, 1991), although detection can be difficult to assess in an interview or a routine neurological visit (Benedict et al., 2002; Fischer et al., 1994; Peyser, Edwards, Poser, & Filskov, 1980). MS-related cognitive decline has been associated with depression (Arnett et al., 1999a; Arnett al., 1999b; Thornton & Raz, 1997). The nature of this association is unclear though, as depression might cause cognitive dysfunction, or it might result from an underlying disease process that also causes the cognitive impairment (Feinstein, 2006). In fact, Feinstein and colleagues (2004) found that MS patients with depression, as compared to patients without depression, tend to have greater lesion load in the prefrontal cortex and anterior temporal lobe in the dominant hemisphere (Feinstein et al., 2004). Other symptoms of the disease may also affect performance on neuropsychological tests. For example, dysarthria or impaired oral agility appears to affect performance on measures dependent on rapid speech, such as the Controlled Oral Word Association Test (Arnett, Smith, Barwick, Benedict, & Ahlstrom, 2008). Similarly, upper extremity motor dysfunction may interfere with performance on cognitive measures that require some aspects of manual dexterity.

With this in mind we constructed regression-based norms for the MACFIMS using data derived from healthy controls. The regression-based continuous norms were used to generate predicted scores for the MS patients, and in turn to calculate *T* scores based on the raw test performance. We then compared these regression-based *T* scores to *T* scores based on published, discrete norms, and evaluated the rates of impairment using each type of norm. Furthermore, we divided MS patients into groups based on neurological functioning, according to performance on tests of oral speed/agility and pegboard placement speed, as well as scores on the Beck Depression Inventory-Fast Screen (BDI-FS; Beck, Steer, & Brown, 2000). We hypothesized that different interpretations would emerge from manual and regression-based standard scores, significantly affecting outcomes on the MACFIMS battery. In addition, we predicted that the severity of neurological abnormality would moderate these findings.

METHOD

Participants

All data were collected in compliance with institutional guidelines. The participants were 395 patients with clinically definite MS (Polman et al., 2005) who were assessed as research volunteers (n = 77), or were referred for clinical assessment (n = 318). The research participants were paid for their participation. Disease course (Lublin & Reingold, 1996) was determined by board-certified neurologists as follows: 294 relapsing-remitting (RR), 84 secondary progressive (SP), 10 progressive-relapsing, and 7 primary progressive (PP). Three hundred eight (308) patients were women and 87 were men. Three hundred sixty-six (366) patients were Caucasian, 23 were African-American, and 6 were classified as "other" reflecting mixed or uncertain heritage. Mean age (±*SD*) was 46.28 ± 8.99 years and mean education was 14.28 ± 2.34 years. Exclusion criteria included history of neurologic disease other than MS, drug or alcohol dependence, and psychiatric disease other than psychological problems attributable to MS. Patients who had experienced relapse or undergone steroid

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

treatment within six weeks prior to participation were also excluded. Expanded Disability Status Scale (EDSS; Kurtzke, 1983) scores within six months were available for 176 patients. Median EDSS score was 3.50 (range 0–7.0).

One hundred healthy adults also participated in the study. These control participants were recruited via advertisement in local, suburban newspapers. All were paid for their participation. All of the control participants were screened for prior neurological and psychiatric illness using a standard screening interview developed in house (Benedict et al., 2006; Parmenter et al., 2007). Most (n = 79) were women, and 89 were Caucasian. On average, these participants were 44.79 years old (±9.43; range: 20–60; skewness: 0.214, standard error = .241; kurtosis: –0.673, standard error = 0.478) and completed 14.47 years of school (±1.72; range: 12–18; skewness: 0.222, standard error 0.241; kurtosis: –.661, standard error = 0.478). The MS patients and healthy controls did not differ significantly on these demographic variables.

Neuropsychological Measures

Neuropsychological testing was conducted by trained assistants and students under the guidance of a board-certified clinical neuropsychologist (RHBB). Patients were administered the MACFIMS battery, as recommended by a consensus panel (Benedict et al., 2002) and recently validated in a large prospective study (Benedict et al., 2006). The battery included: the Judgment of Line Orientation Test (JLO; Benton, Sivan, Hamsher, Varney, & Spreen, 1994), Controlled Oral Word Association Test (COWAT; Benton & Hamsher, 1989), California Verbal Learning Test, 2nd Edition (CVLT2; Delis, Kramer, Kaplan, & Ober, 2000), Brief Visuospatial Memory Test–Revised (BVMTR; Benedict, 1997), Delis-Kaplan Executive Function System (DKEFS) Sorting Test (Delis, Kaplan, & Kramer, 2001), Symbol Digit Modalities Test, oral version (SDMT; Smith, 1982), and a modified Paced Auditory Serial Addition Test (PASAT; Rao, Leo, Bernardin, & Unverzagt, 1991). We included the Total Learning (TL) and Delayed Recall (DR) indices from the CVLT2 and BVMTR and both the Total Correct Sorts (CS) and the Description Score (DS) from the DKEFS Sorting Test. Two trials of the PASAT were administered, one with a 3.0-second interstimulus interval (ISI) and one with a 2.0-second ISI.

Three neurological and psychiatric measures were used to assess the influence of depression, dysarthria, and upper extremity weakness and spasticity. The Beck Depression Inventory–Fast Screen (BDI-FS; Beck, Steer, & Brown, 2000), which has been validated in patients with MS (Benedict, Fishman, McClellan, Bakshi, & Weinstock-Guttman, 2003), was used to quantify depression severity. The Maximum Repetition Rate of Syllables and Multisyllabic Combinations Test (MRR; Kent, Kent, & Rosenbek, 1987; cf Arnett, Smith, Barwick, Benedict, & Ahlstrom, 2008), in which the respondent must repeat phonemes (e.g., "ba–ta–ka") as rapidly as possible for 6 seconds, was used to quantify dysarthria. For this, we recorded the number of triplicate phonemes repeated correctly. The Holyan 9-Hole Peg Test (9HPT; Mathiowetz, Weber, Kashman, & Volland, 1985) requires the participant to insert and then remove nine pegs from holes in a pegboard as quickly as possible. We recorded the average number of seconds required to complete the task twice with each hand, and for all four trials.

Published norms used in the current study include the norms provided with the CVLT2, BVMTR, and the DKEFS manuals. For the JLO and the COWAT, we used the norms provided by Benton, Sivan, Hamsher, Varney, & Spreen (1994). For the PASAT and SDMT, we used the norms provided by Rao (1991).

Statistical Analyses

Throughout the study, the threshold for statistical significance was $p < .05$. For some analyses, effect sizes were calculated with the d statistic. Group differences in age, education, depression, MRR, 9HPT, and performance on the MACFIMS battery measures were evaluated using analyses of variance (ANOVA). Group differences for sex and ethnicity were examined using chi-square analyses.

Demographically adjusted T-scores were calculated for MS patients based on the healthy group's scores. The general procedures are described elsewhere (Heaton et al., 2004; Ivnik et al., 1992; Testa et al., submitted). We first converted the control group's raw scores on each neuropsychological measure to scaled scores ($M = 10$, $SD = 3$) using the cumulative frequency distribution of each measure. This served to normalize all of the test score distributions (see Table 1). We then regressed the resulting scaled scores on age, age-squared, sex, and education, entered en bloc. Plots of regression-standardized residuals predicted values showed that the assumption of homoscedasticity was not violated. Next, we converted the MS participants' raw test scores to scaled scores using the raw-to-scale-score conversions derived from the healthy controls. We then applied the multiple regression equations derived from the healthy controls to compute demographically predicted scores for each MS participant. These predicted scores were then subtracted from each participant's actual scores and the differences were divided by the standard deviation of the controls group's raw residuals for each measure (Table 2). Finally, the resulting values were converted to T scores.

Paired sample t tests were used to evaluate MS participants' T scores based on our regression models and T scores derived from published norms. In addition, MS performance on each neuropsychological measure was classified as either intact ($T > 35$) or impaired ($T \leq 35$) based on T scores derived from each norming method. McNemar tests of dependent proportions were then used to determine if the proportion of participants classified as impaired on each measure differed depending on which norms were used.

For the neurological and psychiatric disability measures, MS patients were assigned to one of three groups based on degree of pathology. For both the MRR and 9HPT, which were normally distributed, patients were classified as follows:

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

Table 1. Raw score to scaled score conversions

			Raw Scores								
			CVLT2		BVMTR					DKEFS	
Scaled Score	COWAT	JLO	Total Learning	Delayed Recall	Total Learning	Delayed Recall	PASAT 3.0	PASAT 2.0	SDMT	Correct Sorts	Description
2		<15	<32		<13		0	0	<43	<5	
3	<20			<6	13	<6		1–14			<20
4	20–21	15	32–35	6	14–15		1–18	15–16	43–45	5	
5	22–23	16–17	36–43	7–8	16	6–7	19–25	17–20	46	6	20–22
6	24–27	18–19	44–46	9	17–19	8	26–29	21–22	47–50	7	23–29
7	28–31	20–21	47–48	10	20–22		30–38	23–25	51–54	8	30–31
8	32–35	22	49–51	11	23–24	9	39–42	26–30	55–56	9	32–34
9	36–39	23	52–53	12	25–26	10	43–48	31–35	57–59		35–37
10	40–43	24–25	54–57	13	27–28	11	49–51	36–39	60–63	10	38–40
11	44–48	26	58–60		29–30		52–54	40–42	64–66	11	41–43
12	49–50	27	61–63	14	31		55–56	43–45	67–69	12	44–47
13	51–53	28	64–66	15	32	12	57–58	46–49	70–72		48–50
14	54–57	29	67–70		33		59	50–52	73–74	13	51–54
15	58–62		71–73		34–35			53–57	75–78	14	55
16	63–71	30	74	16				58–59	79	15	56–59
17	72–77						60		80–87	16	60–63
18	>77		>74		36			60	>87		>63

Impaired = scores >1.5 *SD*s below the control group mean; *Borderline* = scores between 0.5 and 1.5 *SD*s below the control group mean; and *Normal* = scores < 0.5 *SD*s below the control group or better. For the BDI-FS, which was not normally distributed in the MS sample, the following cut-off scores were used, consistent with the test manual: *Normal* = BDI-FS < 3; *Borderline* = BDI-FS 3–8; *Depressed* = BDI-FS > 8.

RESULTS

As shown in Table 3, MS patients reported greater depression than healthy controls on the BDI-FS, $F(1, 493) = 51.46$ $p < .001$. The patients also produced fewer triplicate phonemes on the MRR, $F(1, 376) = 14.52$, $p < .001$] and were slower to complete the 9HPT, $F(1, 461) = 5.78$, $p < .05$. The MS patients performed more poorly than healthy controls on all cognitive measures of the MACFIMS (all p's ≤ 0.05).

Table 4 shows the normal control regression models used to derive *T* scores for the MACFIMS. All models include age, age-squared, sex (male = 1; female = 2), and education. MS raw scores on each of the MACFIMS measures were converted to *T* scores based on the regression-based norms as described earlier. For example, consider a 49-year-old female patient with 14 years of education. Her predicted scaled score on the COWAT is 10.23 [7.456 + 49(–0.264) + 49^2(0.003) + 2(1.921) + 14(0.333)]. Her actual COWAT score of 33 corresponds to a scaled score of 8, according to Table 1. We then divide the difference between her actual and predicted scaled scores (8.0 – 10.23 = –2.23) by the standard deviation of the residual (2.83566), and obtain a *z* score of –0.79, which equals a *T* score of 42.

Table 2. Standard deviation of the residual from healthy controls

	SD Residual
COWAT	2.83566
JLO	2.83793
CVLT2 Total Learning	2.8248
CVLT2 Delayed Recall	2.93344
BVMTR Total Learning	2.70744
BVMTR Delayed Recall	2.37938
PASAT 3.0	2.76076
PASAT 2.0	2.84268
SDMT	2.56298
DKEFS Sorting, sorts	2.60497
DKEFS Sorting, description	2.8354

Comparison of Norms

Table 5 shows mean *T* scores for MS patients calculated using each method. Compared to published norms, the regression-based norms resulted in higher *T* scores on the JLO ($p < .01$), CVLT2-TL ($p < .001$), PASAT 3.0 ($p < .001$), and PASAT 2.0 ($p < .001$). Conversely, regression-based norms resulted in lower *T* scores for the COWAT ($p < .001$), CVLT2-DR ($p < .001$), BVMTR-TL ($p < .001$), BVMTR-DR ($p < .001$), the SDMT ($p < .001$), the DKEFS-CS ($p < .001$), and the DKEFS-DS ($p < .001$).

Performance on each measure of the MACFIMS was classified as impaired or intact based on *T* scores derived from each set of norms. The proportion of intact to impaired MS

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

Table 3. MS patients compared to healthy controls

	MS (N = 395) (308 females, 87 males)		Controls (N = 100) (79 females, 21 males)			
	Mean	*SD*	Mean	*SD*	*p* value	*d*'
Age	46.28	8.99	44.79	9.43	*ns*	0.16
Education	14.28	2.34	14.47	1.72	*ns*	0.09
BDI-FS	3.39	3.40	0.90	1.40	<.001	0.96
MRR3	1.37	0.52	1.61	0.39	<.001	0.52
9-HPT	24.78	21.70	18.93	3.39	<.05	0.38
COWAT	34.78	11.69	41.51	11.70	<.001	0.58
JLO	22.30	5.50	24.11	3.90	<.01	0.38
CVLT2 Total Learning	48.80	10.56	55.79	9.35	<.001	‘0.70
CVLT2 Long Delay	10.08	3.22	12.27	2.55	<.001	0.75
BVMTR Total Learning	20.92	7.17	26.68	5.56	<.001	‘0.90
BVMTR Long Delay	8.28	2.82	10.34	1.68	<.001	0.89
PASAT 3.0 ISI	40.04	13.36	46.30	12.19	<.001	0.48
PASAT 2.0 ISI	29.38	10.67	36.21	11.24	<.001	0.62
SDMT	49.53	13.20	61.40	9.31	<.001	1.04
DKEFS Sorting, Correct Sorts	9.48	2.45	10.28	2.33	<.01	0.33
DKEFS Sorting, Description Score	35.59	10.54	38.61	10.07	<.05	0.29

patients was calculated for each measure. As seen in Table 6, the regression-based norms resulted in significantly more patients being classified as impaired on the CVLT2-DR ($p < .05$), BVMTR-TL ($p < .001$), BVMTR-DR ($p < .001$), SDMT ($p < .001$), DKEFS-CS ($p < .001$), and DKEFS-DS ($p < .001$). On the other hand, the published, discrete norms classified significantly more patients as impaired on the CVLT2-TL ($p < 0.001$), PASAT 3.0 ($p < 0.001$), and PASAT 2.0 ($p < 0.001$). Both norms classified similar numbers of patients as impaired on the COWAT and JLO.

Comparisons of Groups Based on Neurologic Symptoms

On the MRR, 119 patients were classified as normal (group 1), 89 were classified as borderline (group 2), and 93 were classified as impaired (group 3). These subgroups did not differ on age, education, sex, or race. A multivariate analysis of variance (MANOVA) on MACFIMS T scores derived from the regression-based norms revealed significant group differences, $F(22, 576) = 1.74$, $p = .020$. Follow-up analyses were conducted and, as seen in Table 7, significant group differences were found on the CVLT2-TL, $F(2, 298) = 4.02$, $p = .019$, CVLT2-DR, $F(2, 298) = 3.26$, $p = .040$, PASAT 3.0, $F(2, 298) = 3.66$, $p = .027$, and SDMT, $F(2, 298) = 11.991$, $p < .001$. On the CVLT2-TL and SDMT, group 1 performed better than groups 2 and 3. On the PASAT 3.0, group 1 performed better than group 3 only. On the CVLT2-DR, group 1 performed better than groups 2 and 3, although these comparisons only approached significance (p's < .09).

On the 9HPT, there were 135 patients classified as normal (group 1), 94 as borderline (group 2), and 153 as impaired (group 3). These subgroups differed in age, $F(2, 379) = 20.68$, $p < .001$, with group 1 being significantly younger (42.45 ± 8.26 years) than groups 2 (47.52 ± 9.18 years) and 3 (48.80 ± 8.64 years). The groups did not significantly differ in education, sex, or race. A MANOVA revealed significant group differences, $F(22, 738) = 7.280$, $p < .001$, on the MACFIMS. As seen in Table 8, the groups differed on all measures (all p's < .05). Group 1 performed significantly better than group 3 on all measures (all p's < .05). Additionally, group 1 performed significantly better than group 2 on the COWAT ($p < .05$), CVLT2-TL ($p < .05$), CVLT2-DR ($p < .05$), PASAT 2.0 ($p < .01$), and SDMT ($p < .01$). Group 2 performed significantly better than group 3 on the JLO ($p < .05$), CVLT2-TL ($p < .01$), CVLT2-DR ($p < .05$), PASAT 3.0 ($p < .05$), and SDMT ($p < .001$). When a multivariate analysis of covariance (MANCOVAs) was conducted, with age used as a covariate, significant group differences were again found, $F(22, 736) = 7.64$, $p < .001$, and group differences were found for all measures (all p's < .01).

On the BDI-FS, 200 patients were normal (group 1), 159 patients were borderline (group 2), and 36 patients were depressed (group 3). These groups did not differ in age, education, or sex. However, group 3 included fewer African-Americans and more patients of "other" racial background than groups 1 and 2, χ^2 (4) = 14.33, $p = .006$. However, as shown in Table 9, a MANOVA revealed no group differences on the MACFIMS.

DISCUSSION

The purposes of this study were three-fold: (1) establish regression-based norms for use with the MACFIMS, (2) compare these norms with the traditional manual-based norms published for each test, and (3) assess the relationship between cognitive performance and neurological and psychiatric factors.

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

Table 4. Final regression models for MACFIMS measures

Measure	Predictor	B	Standard Error B	T	Standardized B	Total R square
COWAT	(constant)	7.456	5.928	1.258		
	age	–0.264	0.246	–1.072	–0.823	
	age^2	0.003	0.003	0.995	0.761	
	sex	1.921	0.260	2.649	0.260	
	education	0.333	0.174	1.909	0.187	0.119
JLO	(constant)	7.303	5.932	1.231		
	age	0.292	0.246	1.185	0.92	
	age^2	–0.004	0.003	–1.25	–0.967	
	sex	–2.16	0.726	–2.977	–0.296	
	education	0.078	0.175	0.449	0.045	0.099
CVLT2 Total Learning	(constant)	5.533	5.905	0.937		
	age	–0.038	0.245	–0.154	–0.119	
	age^2	0	0.003	0.078	0.06	
	sex	2.451	0.722	3.393	0.334	
	education	0.093	0.174	0.534	0.053	0.114
CVLT2 Delayed Recall	(constant)	1.648	6.132	0.269		
	age	0.245	0.255	0.962	0.758	
	age^2	–0.003	0.003	–1.077	–0.846	
	sex	1.661	0.75	2.215	0.224	
	education	0.073	0.18	0.404	0.041	0.07
BVMTR Total Learning	(constant)	–3.216	5.66	–0.568		
	age	0.437	0.235	1.862	1.368	
	age^2	–0.006	0.003	–2.303	–1.685	
	sex	0.844	0.692	1.22	0.115	
	education	0.391	0.167	2.347	0.22	0.194
BVMTR Delayed Recall	(constant)	5.333	4.974	1.072		
	age	0.279	0.206	1.353	1.042	
	age^2	–0.004	0.002	–1.747	–1.34	
	sex	0.458	0.608	0.752	0.074	
	education	0.01	0.146	0.07	0.007	0.114
PASAT 3.0	(constant)	–0.373	5.771	–0.065		
	age	0.188	0.24	0.784	0.597	
	age^2	–0.003	0.003	–0.977	–0.742	
	sex	–0.247	0.706	–0.35	–0.034	
	education	0.568	0.17	3.344	0.326	0.132
PASAT 2.0	(constant)	6.238	5.942	1.05		
	age	0.031	0.247	0.126	0.098	
	age^2	–0.001	0.003	–0.399	–0.308	
	sex	–0.475	0.727	–0.654	–0.065	
	education	0.391	0.175	2.238	0.222	0.104
SDMT	(constant)	6.838	5.358	1.276		
	age	0.199	0.222	0.895	0.619	
	age^2	–0.004	0.003	–1.647	–1.135	
	sex	0.773	0.655	1.179	0.104	
	education	0.135	0.158	0.858	0.076	0.285
DKEFS Sorting Correct Sorts	(constant)	–3.316	5.445	–0.609		
	age	0.231	0.226	1.02	0.725	
	age^2	–0.004	0.003	–1.416	–1.004	
	sex	0.794	0.108	1.191	0.108	
	education	0.668	0.16	4.166	0.379	0.244
DKEFS Sorting Description	(constant)	3.485	5.927	0.588		
	age	–0.005	0.246	–0.02	–0.016	
	age^2	0.000	0.003	–0.213	–0.164	
	sex	0.756	0.725	1.043	0.103	
	education	0.459	0.174	2.63	0.259	0.113

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

Table 5. Comparison of MS patients' mean *T* scores for the MACFIMS calculated from published norms versus regression-based norms

	Published Norms		Regression-based Norms			
	M	SD	M	SD	*t*(*df*)	*d'*
COWAT	44.67	11.07	43.58	10.20	3.86 (394)***	0.10
JLO	47.91	10.89	48.69	11.65	2.97 (394)**	0.07
CVLT2 Total Learning	48.34	11.79	53.49	11.22	19.51 (394)***	0.45
CVLT2 Delayed Recall	43.95	11.44	40.77	10.31	12.83 (394)**	0.29
BVMTR Total Learning	43.37	12.49	36.62	12.44	21.35 (394)***	0.54
BVMTR Delayed Recall	45.64	12.90	37.68	12.90	32.73 (394)***	0.62
PASAT 3.0	40.35	13.18	47.03	10.73	16.10 (394)***	0.56
PASAT 2.0	41.31	10.64	42.76	9.80	5.33 (394)***	0.14
SDMT	41.51	12.09	34.19	13.35	23.52 (394)***	0.57
DKEFS Sorting sorts	51.33	10.59	48.47	12.25	5.74 (394)***	0.25
DKEFS Sorting description	51.96	10.96	42.44	10.99	21.00 (394)***	0.87

**$p < .01$
***$p < .001$.

To our knowledge, this is the first application of regression-based normative techniques to the MACFIMS. This approach enables one to compare performance across tests directly, because the norms are derived from a uniform data set (i.e., they are co-normed), rather than from different standardization samples for each test. In the regression analyses we included the traditional demographic variables typically used in regression-based norms, such as, age, age^2, sex, and education. Including a term for age-squared allowed us to consider the nonlinear relationship between age and cognition. The regression-based approach to norms development enables one to (a) account for demographic influences on test performance, and (b) use the entire normative sample rather than divide it into smaller subgroups for the computation of age- or education-stratified means and standard deviations. Because the control group was demographically matched to the larger clinical sample, and because the clinical sample is representative of MS clinical patients, these norms may be applicable in other regions of the USA. A challenge to the regression approach, however, is determining which variables are reasonable predictors. There is ample evidence that age, education, and, to a lesser extent, sex, correlate with neuropsychological test performance. Other variables can be included, but it is important to consider whether they make theoretical sense. For example, height might be a statistically significant predictor, but there is little theoretical basis for including it as a predictor in regression-based normative equations. In the future, we may pursue similar regression-based approaches within the MS sample that may enable us to account for the influence of peripheral or neurological factors, such as MRR and 9HPT.

Table 6. Number of impaired patients classified by discrete and regression-based norms

	Discrete Norms	Regression-based norms
COWAT	78 (19.75%)	84 (21.27%)
JLO	53 (13.42%)	57 (14.43%)
CVLT2 Total Learning***	50 (12.66%)	12 (3.04%)
CVLT2 Delayed Recall*	114 (28.86%)	129 (32.66%)
BVMTR Total Learning***	120 (30.38%)	191 (48.35%)
BVMTR Delayed Recall***	103 (26.08%)	171 (43.29%)
PASAT 3.0***	153 (38.75%)	46 (11.65%)
PASAT 2.0***	118 (29.87%)	91 (23.04%)
SDMT***	134 (33.92%)	216 (54.68%)
DKEFS Sorting- sorts***	17 (4.30%)	45 (11.39%)
DKEFS Sorting- description***	21 (5.32%)	105 (26.58%)

*p<0.05
***p<0.001

Practical issues should also be considered when the clinician is determining whether to use regression-based norms. As Van der Elst, Van Boxtel, Van Breukelen, & Jolles (2006a; 2006b) observed, many clinicians are unfamiliar with such norms and might find them cumbersome. These authors recommend implementing a spreadsheet on a computer to help overcome this. On the other hand, an advantage of regression-based norms is that the clinician no longer has to compile separate norms from several sources to evaluate the patient.

Although several authors advocate using regression-based norms (Crawford & Howell, 1998; Heaton, Avitable, Grant, & Matthews, 1999; Zachary & Gorsuch, 1985), others (Reitan & Wolfson, 1995; 1997) argue that demographic variables predict cognitive test performance in healthy, neurologically intact individuals, but that these relationships are uncoupled by brain damage or dysfunction. Fasteneau (1998) also cautioned against using regression-based norms, based on the view that such models can distort findings. For example, it was reported that such models penalized highly educated adults on the Trail Making Test, Boston Naming

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

Table 7. MRR3 Group Performance on the MACFIMS (regression-based T scores)

	MRR3 Group 1/ Normal (N = 119)	MRR3 Group 2/ Borderline (N = 89)	MRR3 Group 3/ Impaired (N = 93)	p-value
COWAT	44.43 (10.47)	43.21 (9.07)	43.41 (10.66)	ns
JLO	48.92 (11.57)	47.51 (11.45)	47.86 (11.86)	ns
CVLT2 Total Recall	56.07 (11.06)	52.35 (11.27)	52.28 (11.45)	<0.05
CVLT2 Delayed Recall	42.69 (10.37)	39.66 (10.11)	39.60 (9.99)	<0.05
BVMTR Total Recall	37.75 (13.03)	35.36 (12.95)	36.70 (11.08)	ns
BVMTR Delayed Recall	39.06 (12.71)	37.47 (12.69)	36.62 (12.69)	ns
PASAT 3.0	49.06 (11.97)	45.93 (10.39)	45.48 (8.68)	<0.05
PASAT 2.0	43.92 (9.12)	41.65 (10.28)	42.27 (8.91)	ns
SDMT	39.13 (13.68)	33.72 (13.83)	30.40 (11.64)	<0.001
DKEFS Correct Sorts	48.84 (11.48)	46.80 (15.07)	48.09 (10.26)	ns
DKEFS Description Score	43.32 (10.37)	41.73 (12.23)	41.77 (9.32)	ns

Test, and Wisconsin Card Sorting Test (WCST), while overcorrecting for age on the WCST (Fasteneau, 1998). These arguments have been challenged (Shuttleworth-Jordan, 1997; Vanderploeg, Axelrod, Sherer, Scott, & Adams, 1997) on both methodological and theoretical grounds. Indeed, while Heaton, Avitable, Grant, and Matthews (1999) agreed that regression-based norms can be misleading, they noted that this is a result of an inappropriate sample from which the models are derived and is not inherent to these types of norms. However, regression-based norms may not be appropriate if the referral question focuses on real-world functioning. In a recent paper, Silverberg and Millis (2009) argue that, while demographically adjusted norms can determine if a patient has deteriorated from baseline (i.e., impairment), norms that do *not* account for demographic variables better predict a person's current functional abilities or lack thereof (i.e., deficiency), such as the ability to drive or live independently. Thus, the clinician will need to decide which type of norm is most appropriate, either those that account for demographic variables or those that do not, depending on the referral question.

Our findings show that regression-based norms yield significantly different *T* scores than discrete norms published for each test. As a result, *T* scores were significantly altered for all test variables: COWAT, JLO, CVLT2-TL, CVLT2-DR, BVMTR-TL, BVMTR-DR, PASAT 3.0, PASAT 2.0, SDMT, DKEFS-CS, and the DKEFS-DS. This has important clinical implications because whether or not a person is diagnosed with a cognitive disorder or dementia can vary depending, in part, on which norms are used. Our results suggest that published norms may be inadequate for interpreting performance on the MACFIMS. Compared to our regression-based adjustments, published norms resulted in significantly lower rates of impairment for MS patients on the CVLT2-DR, the BVMTR, SDMT, and the DKEFS Sorting Test. Most of the published norms that we used were stratified by age and education, but accounted for these as categorical rather than continuous variables. Consequently, discrete norms might distort the effect of these variables, such as the impact of education on the DKEFS Sorting Test.

It should be borne in mind that our study was not designed to determine which norm method is most valid. The manual norms employed in this study were collected by different researchers at different times. Some of these data were published in 1991 (e.g., for PASAT and SDMT), while others were published more recently (e.g., DKEFS). One of our goals was to collect and derive new normative data based on a single sample, so that comparisons across tests would be

Table 8. 9HPT Group Performance on the MACFIMS (regression-based T scores)

	9HPT Group 1/ Normal (N = 135)	9HPT Group 2/ Borderline (N = 94)	9HPT Group 3/ Impaired (N = 153)	p-value
COWAT	46.70 (10.35)	43.19 (8.99)	41.01 (9.77)	<0.001
JLO	50.56 (10.90)	50.60 (11.55)	46.23 (11.73)	<0.01
CVLT2 Total Recall	57.77 (10.63)	53.92 (10.93)	49.72 (10.40)	<0.001
CVLT2 Delayed Recall	44.33 (10.48)	41.01 (11.00)	37.69 (8.82)	<0.001
BVMTR Total Recall	39.50 (11.86)	37.08 (12.31)	34.24 (12.79)	<0.01
BVMTR Delayed Recall	40.74 (11.78)	38.76 (12.14)	34.94 (13.70)	<0.001
PASAT 3.0	49.10 (10.61)	47.63 (11.38)	44.67 (9.99)	<0.01
PASAT 2.0	46.07 (9.16)	42.15 (9.16)	40.33 (9.81)	<0.001
SDMT	41.52 (11.97)	35.96 (11.95)	26.89 (11.51)	<0.001
DKEFS Correct Sorts	50.15 (13.41)	49.45 (11.48)	46.73 (11.36)	<0.05
DKEFS Description Score	45.92 (11.52)	42.75 (10.49)	39.59 (9.94)	<0.001

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

Table 9. BDI-FS Group Performance on the MACFIMS (regression-based T scores)

	BDI-FS Group 1/ Normal (N = 200)	BDI-FS Group 2/ Borderline (N = 159)	BDI-FS Group 3/ Depressed (N = 36)
COWAT	44.33 (10.65)	42.81 (9.63)	42.87 (10.07)
JLO	49.23 (11.51)	47.43 (11.83)	51.30 (11.23)
CVLT2 Total Recall	54.90 (10.73)	52.41 (11.50)	50.36 (11.77)
CVLT2 Delayed Recall	41.89 (9.76)	39.91 (11.07)	38.35 (9.21)
BVMTR Total Recall	37.60 (12.45)	36.05 (12.10)	33.70 (13.62)
BVMTR Delayed Recall	39.10 (13.14)	36.45 (12.27)	35.23 (13.66)
PASAT 3.0	47.71 (11.06)	46.38 (10.29)	46.10 (10.77)
PASAT 2.0	43.20 (9.65)	42.20 (9.45)	42.84 (12.09)
SDMT	36.13 (13.48)	32.33 (12.95)	31.68 (13.14)
DKEFS Correct Sorts	48.90 (12.23)	47.65 (12.58)	49.63 (10.93)
DKEFS Description Score	42.58 (10.70)	42.14 (11.58)	43.01 (10.09)

more valid. As we do not have a gold standard for comparison, we cannot determine which norms are most valid. However, we do note a few observations that would seem to support the regression-based approach. First, only 4 to 5% of patients are impaired on the DKEFS using the manual norms, but we and other researchers have shown previously that the DKEFS is as sensitive as the Wisconsin Card Sorting Test in MS (Beatty & Monson, 1996; Parmenter et al, 2007). For reasons that are not clear, the manual norms for this test appear to be generous, possibly compromising the test's sensitivity.

We also examined how neurological symptoms other than cognitive impairment might affect neuropsychological functioning. As mentioned previously, we hypothesized that patients with dysarthria, as measured by reduced performance on the maximum rate repetition test, would have reduced performance on measures reliant on speech (i.e., the COWAT, CVLT2, SDMT, and the PASAT). Additionally, we hypothesized that the severity of neurological abnormality, such as impaired performance on the MMR3 and 9HPT, would moderate our findings. For example, patients with poor manual dexterity, as measured by the 9HPT, would be expected to perform more poorly on measures dependent on manual dexterity, such as the BVMTR. Consistent with this, we found that patients impaired on the MRR3 performed significantly lower on several measures reliant on speech, such as the CVLT2, PASAT 3.0, and the SDMT (although no significant differences were found on the COWAT or PASAT 2.0). These findings suggest that neurological symptoms separate from cognition may influence performance on neuropsychological measures. This was not seen for other MS symptoms, however. MS patients with reduced manual dexterity not only performed more poorly on measures dependent on this function, but they performed more poorly on all measures of the MACFIMS. Conversely, no differences on the MACFIMS were found for patients grouped according to depression. Reduced performance on the 9HPT may represent more advanced disease, related to global cognitive impairment. However, the lack of findings related to severity of depression is uncertain, as these data are in direct contrast to previous findings, despite similar methods used to define depression (Arnett et al., 1999a, 1999b; Thornton & Raz, 1997).

One limitation of our study is the relatively small number of healthy controls on which our regression models were based. Even though one advantage of using regression-based norms is that fewer people are needed for the normative sample, a larger number would provide the models derived for this analysis with greater stability. Thus, deriving models from a larger normative sample may be a useful future endeavor.

In sum, our findings suggest that existing discrete norms might overlook impairment in patients with MS, which could prevent them from receiving needed assistance. Additionally, by controlling the effects of demographic variables, we can better appreciate how other MS symptoms might contribute to neuropsychological performance, such as the effect of dysarthria on tests that require oral responses. Thus, these norms provide an alternative when using the MACFIMS to evaluate a patient's cognitive functioning and to investigate how MS affects cognition.

ACKNOWLEDGMENTS

There were no sources of financial support for this article. The authors have no financial or other relationships that could be interpreted as a conflict of interest with regard to this article.

REFERENCES

Arnett, P.A., Higginson, C.I., Voss, W.D., Bender, W.I., Wurst, J.M., & Tippin, J.M. (1999a). Depression in multiple sclerosis: Relationship to working memory capacity. *Neuropsychology*, *13*(4), 546–56.

Arnett, P.A., Higginson, C.I., Voss, W.D., Wright, B., Bender, W.I., Wurst, J.M., & Tippin, J.M. (1999b). Depressed mood in multiple sclerosis: Relationship to capacity-demanding memory and attentional functioning. *Neuropsychology*, *13*(3), 434–446.

Arnett, P.A., Smith, M.M., Barwick, F.H., Benedict, R.H.B., & Ahlstrom, B.P. (2008). Oralmotor slowing in multiple sclerosis: Relationship to neuropsychological tasks requiring an oral

https://doi.org/10.1017/S1355617709990750 Published online by Cambridge University Press

response. *Journal of the International Neuropsychological Society*, *14*, 454–462.

Beatty, W.W., & Monson, N. (1996). Problem solving by patients with multiple sclerosis: Comparison of performance on the Wisconsin and California Card Sorting tests. *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 Antonio, TX: Psychological Corporation.

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

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., Fischer, J.S., Archibald, C.J., Arnett, P.A., Beatty, W.W., Bobholz, J., et al. (2002). Minimal neuropsychological assessment of MS patients: A consensus approach. *The Clinical Neuropsychologist*, *16*(3), 381–397.

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

Benton, A.L., Sivan, A.B., Hamsher, K., Varney, N.R., & Spreen, O. (1994). *Contributions to Neuropsychological Assessment*. Second ed. New York: Oxford University Press.

Crawford, J.R., & Allan, K.M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, *11*(2), 192–197.

Crawford, J.R., & Howell, D.C. (1998). Regression equations in clinical neuropsychology: An evaluation of statistical methods for comparing predicted and obtained scores. *Journal of Clinical and Experimental Neuropsychology*, *20*(5), 755–762.

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

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Function System*. San Antonio, Texas: Psychological Corporation.

Fastenau, P.S. (1998). Validity of regression-based norms: An empirical test of the comprehensive norms with older adults. *Journal of Clinical and Experimental Neuropsychology*, *20*(6), 906–916.

Feinstein, A. (2006). Mood disorders in multiple sclerosis and the effects on cognition. *Journal of the Neurological Sciences*, *245*, 63–66.

Feinstein, A., Roy, P., Lobaugh, N., Feinstein, K., O'Connor, P., & Black, S. (2004). Structural brain abnormalities in multiple sclerosis patients with major depression. *Neurology*, *62*, 586–590.

Fischer, J.S., Foley, F.W., Aikens, J.E., Ericson, G.D., Rao, S.M., & Shindell, S. (1994). What do we really know about cognitive dysfunction, affective disorders, and stress in multiple slcerosis? A practitioner's guide. *Journal of Neurologic Rehabilitation*, *8*, 151–164.

Heaton, R.K., Avitable, N., Grant, I., & Matthews, C.G. (1999). Further cross validation of regression-based neuropsychological norms with an update for the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology*, *21*(4), 572–582.

Heaton, R.K., Ryan, L., Grant, I., & Matthews, C.G. (1996). Demographic influences on neuropsychological test performance. In I. Grant & K. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders*. New York: Oxford University Press.

Kent, R.D., Kent, J.F., & Rosenbek, J.C. (1987). Maximum performance tests of speech production. *Journal of Speech and Hearing Disorders*, *52*, 367–387.

Klein, C., Forester, F., & Hartnegg, K. (2007). Regression-based developmental models exemplified for Wisconsin Card Sorting Test parameters: Statistics and software for individual predictions. *Journal of Clinical and Experimental Neuropsychology*, *29*(1), 25–35.

Kurtzke, J.F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Annals of Neurology*, *13*, 227–231.

Leckliter, I.N., & Matarazzo, J.D. (1989). The influence of age, education, IQ, gender, and alcohol abuse on Halstead-Reitan neuropsychological test battery performance. *Journal of Clinical Psychology*, *45*(4), 484–512.

Lublin, F.D., & Reingold, S.C. (1996). Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. [see comment]. *Neurology*, *46*, 907–911.

Mathiowetz, V., Weber, K., Kashman, N., & Volland, G. (1985). Adult norms for the nine hole peg test of finger dexterity. *The Occupational Therapy Journal of Research*, *5*, 24–37.

McDonald, W.I., Compston, A., Edan, G., Goodkin, D.E., Hartung, H., Lublin, F., McFarland, H.F., Paty, D.W., Polman, C.H., Reingold, S.C., Sandberg-Wollheim, M., Sibley, W.A., Thompson, A., van der Noort, S., Weinshenker, B.Y., & Wolinsky, J.S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, *50*, 121–127.

Parmenter, B., Zivadinov, R., Kerenyi, L., Gavett, R., Weinstock-Guttman, B., Dwyer, M., et al. (2007). Validity of the Wisconsin Card Sorting and Delis-Kaplan Executive Function System (DKEFS) sorting tests in multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *29*, 215–223.

Peyser, J.M., Edwards, K.R., Poser, C.M., Filskov, S.B. (1980). Cognitive function in patients with multiple sclerosis. *Archives of Neurology*, *37*, 577–579.

Polman, C.H., Reingold, S.C., Edan, G., Filippi, M., Hartung, H.P., Kappos, L., et al. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria." *Annals of Neurology*, *58*, 840–846.

Rao, S.M. (1991). *A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis*. New York: National MS Society.

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

Reitan, R.M., & Wolfson, D. (1995). Influence of age and education on neuropsychological test results. *The Clinical Neuropsychologist*, *9*(2), 151–158.

Reitan, R.M. & Wolfson, D. (1997). The influence of age and education on neuropsychological performances of persons with mild head injuries. *Applied Neuropsychology*, *4*(1), 16–33.

Schretlen, D.J., Cascella, N.G., Meyer, S.M., Kingery, L.R., Testa, S.M., Munro, C.A., Pulver, A.E., Rivkin, P., Rao, V.A., Diaz-Asper, C.M., Dickerson, F.B., Yolken, R.H., & Pearlson, G.D. (2007). Neuropsychological functioning in bipolar disorder and schizophrenia. *Biol Psychiatry*, *62*, 179–186.

Sherrill-Pattison, S., Donders, J., & Thompson, E. (2000). Influence of demographic variables on neuropsychological test performance after traumatic brain injury. *The Clinical Neuropsychologist*, *14*(4), 496–503.

Shuttleworth-Jordan, A.B. (1997). Age and education effects on brain-damaged subjects: "Negative" findings revisited. *The Clinical Neuropsychologist*, *11*(2), 205–209.

Silverberg, N.D., & Millis, S.R. (2009). Impairment versus deficiency in neuropsychological assessment: Implications for ecological validity. *Journal of the International Neuropsychological Society*, *15*, 94–102.

Smith, A. (1982). *Symbol digit modalities test: Manual*. Los Angeles: Western Psychological Services.

Testa, S.M., Winicki, J. M, Pearlson, G.D., Gordon, B., & Schretlen, D.J. (submitted). Accounting for estimated IQ in neuropsychological test performance with regression-based norms.

Thornton, A., & Raz, N. (1997). Memory impairment in multiple sclerosis: A quantitative review. *Neuropsychology*, *11*(3), 357–366.

Van Breukelen, G.J.P., & Vlaeyen, J.W.S. (2005). Norming clinical questionnaires with multiple regression: The Pain Cognition List. *Psychological Assessment*, *17*(3), 336–344.

Van der Elst, W., Van Boxtel, M.P.J., Van Breukelen, G.J.P., & Jolles, J. (2005). Rey's verbal learning test: Normative data for 1855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of transportation. *Journal of the International Neuropsychological Society, 11*, 290–302.

Van der Elst, W., Van Boxtel, M.P.J., Van Breukelen, G.J.P., & Jolles, J. (2006a). The Stroop Color-Word Test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, *13*(1), 62–79.

Van der Elst, W., Van Boxtel, M.P.J., Van Breukelen, G.J.P., & Jolles, J. (2006b). The Concept Shifting Test: Adult normative data. *Psychological Assessment, 18*(4), 424–432.

Vanderploeg, R.D., Axelrod, B.N., Sherer, M., Scott, J., & Adams, R.L. (1997). The importance of demographic adjustments on neuropsychological test performance: A response to Reitan and Wolfson (1995). *The Clinical Neuropsychologist*, *11*(2), 211–217.

Zachary, R.A., & Gorsuch, R.L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41*(1), 86–94.