

Determining the Presence of Reliable Change over Time in Multiple Sclerosis

Evidence from the PASAT, Adjusting-PSAT, and Stroop Test

Suzanne L. Barker-Collo, PhD; Suzanne C. Purdy, PhD

*People with multiple sclerosis (MS) often undergo repeated assessments. Methods for determining whether an individual's change in test results over time is reliable require further study. A sample of individuals with MS (N = 52) was assessed at baseline and at 6-month follow-up using the Paced Auditory Serial Addition Test (PASAT), Simple Adjusting-Paced Serial Addition Test (A-PSAT), and Victoria Stroop test. Two methods for determining the reliability of an individual's change over time were examined. The Reliable Change Index (RCI) identified few individuals who declined reliably between baseline and 6-month follow-up. The standard deviation (SD) method of calculation resulted in reliable declines for a small number of individuals on most measures. Use of the SD method resulted in a larger number of individuals who improved reliably. Responsiveness of individuals to treatment effects can be lost in group analyses. The data presented here provide clinicians with an approach for determining whether an individual's change over time on commonly used neuropsychological tests reflects reliable improvement or decline. *Int J MS Care*. 2013;15:170–178.*

The presence of cognitive dysfunction in multiple sclerosis (MS) has been well established.¹ At every stage of the disease, even prior to diagnosis, between 40% and 65% of individuals experience cognitive impairment.^{2,3} The most commonly reported areas of cognitive dysfunction in MS are attention, concentration, memory, information processing speed, and some aspects of executive functioning.^{2,4} While patients' neuropsychological strengths and weaknesses vary depending on the brain region or regions affected—resulting in a heterogeneous MS “profile” across patients—information processing speed is a key area of deficit.^{5–10}

Repeated cognitive assessment—particularly of information processing speed—is a common method of tracking disease progression, whether or not treatments such as beta-interferons are being used.¹¹ In fact, a measure of information processing speed, the abbreviated form of the Paced Auditory Serial Addition Test

(PASAT), is the only cognitive measure included in the Multiple Sclerosis Functional Composite (MSFC) index developed by the US National Multiple Sclerosis Society Task Force on Clinical Outcomes.^{12,13} The importance of accurately tracking change in cognitive functioning in MS is indicated by findings suggesting that cognitive functioning makes a significant and independent contribution to activities of daily living, quality of life, and work and social functioning.^{2,14} Yet the literature emphasizes that clinicians should examine change in performance over time not only in terms of statistical significance, but also in terms of clinical significance. In recent years, several techniques have been developed for the purpose of determining clinically significant levels of change; the two most commonly used are the standard deviation (SD) method and the Reliable Change Index (RCI).^{15,16}

One of the issues arising when administering tests of cognitive functioning repeatedly over time is the presence of test-retest variability and practice effects. This makes it difficult to interpret longitudinal neuropsychological test data, as changes in scores cannot be assumed to reflect true change in cognitive status. The SD, RCI, and other methods have been proposed to account for variability and practice effects. Although these methods

From the Department of Psychology, Faculty of Science, University of Auckland, Auckland, New Zealand. Correspondence: Suzanne L. Barker-Collo, PhD, Department of Psychology, Tamaki Campus, The University of Auckland, Private Bag 92019, Auckland, New Zealand; e-mail: s.barker-collo@auckland.ac.nz.

DOI: 10.7224/1537-2073.2013-007

© 2013 Consortium of Multiple Sclerosis Centers.

have been examined extensively in other disease populations,¹⁷⁻¹⁹ only a few studies have addressed them in people with MS.^{1,20,21}

The objective of this study was to examine change in performance of people with MS on commonly administered tests of information processing speed using two published methods available to clinicians for determining clinically significant change (the SD and RCI methods) with a 6-month test-retest interval. The tests included were those commonly used in longitudinal evaluations of change in this population. Data were derived for the sample as a whole, as well as for relapsing-remitting (RRMS) and primary progressive (PPMS) subgroups.

Methods

Participants

Participants were 52 community-dwelling individuals with a primary diagnosis of MS. They included 16 men (30.8%) and 36 women (69.2%) who ranged in age from 27 to 78 years, with a mean age of 53.2 years. Amount of education completed ranged from 9 to 21 years, with a mean of 13.1 years. The majority of participants were of European ancestry ($n = 50$, 96.2%). The remaining two participants (3.8%) self-identified as being of Maori ancestry. Each of these two participants had one biological parent of Maori ancestry and one parent of European ancestry. Potential participants were excluded from the study if they had 1) another medical condition that could affect performance or change in performance over time (eg, major depression, diabetes, psychosis, dementia), or 2) physical or sensory impairments that would affect their ability to complete the tasks (eg, color blindness, aphasia, inability to press computer keys). With regard to use of substances that had the potential to affect the findings, 7 (13.5%) participants were taking fish oil supplements and 6 (11.5%) were taking painkillers. In both cases, these were taken on a regular basis and consistently throughout the study. In addition, 2 individuals (3.8%) had undergone interferon treatment within the past 5 years, although neither received such treatment at any time during the study.

With regard to MS type as indicated in clinical records, 29 participants (55.8%) had RRMS, reporting an average of 7.2 relapses. The remaining 23 participants (44.2%) had PPMS. None of the participants in the

RRMS group reported experiencing a relapse during this study. Participants reported that they had received their MS diagnoses an average of 13 years prior to the initial assessment (range, 6 months to 36 years) and had experienced symptoms that they currently associated with MS for an average of 7.8 years before receiving the diagnosis.

Measures

Expanded Disability Status Scale

The Expanded Disability Status Scale (EDSS)²² evaluates disability numerically. A patient is assessed on the EDSS by a neurologist according to observed signs and symptoms. In the present study, all EDSS scores were obtained from patient medical records made by the attending neurologist. An ordinal clinical rating scale, the EDSS ranges from 0 (normal neurologic examination) to 10 (death due to MS) in half-point increments. The patient is evaluated on eight Functional Systems: pyramidal, cerebellar, brainstem, sensory, bowel and bladder, visual, cerebral, and other. Each Functional System scale ranges from 0 to 5 or 6. Interrater reproducibility is adequate for group comparison studies (Spearman $r = 0.78$), while intrarater reproducibility is more variable (Spearman $r = 0.62-0.94$).²³ Convergent and discriminant validity for the EDSS is supported. Both test-retest reliability and interrater agreement have varied considerably across studies, with some finding high values and others unacceptably low figures.²⁴

Paced Auditory Serial Addition Test

The PASAT²⁵ is an audiotaped measure that presents a list of single-digit numbers to the subject. The subject must add each number to the preceding number and state the result. Over four trials, the speed of presentation increases from an interstimulus interval (ISI) of 2.4 seconds to 2.0 seconds to 1.6 seconds to 1.2 seconds; a total of 61 numbers are presented in each trial. For this study, only the 2.4- and 2.0-second intervals were used. Total scores on each trial were the total number of correct responses produced, with standard scores calculated based on age-adjusted normative data.

Simple Adjusting-Paced Serial Addition Test

The Simple Adjusting-Paced Serial Addition Test (A-PSAT)²⁶ is a computerized assessment in which a series of digits from 1 to 9 is presented on the screen with a visual mask following each digit to prevent afterimages. Stimuli appear for 350 ms. For each trial, participants are required to add the two preceding num-

bers and state the sum aloud, and the examiner enters the responses using a keyboard. The task begins with 20 practice trials, followed by 200 scored trials. The A-PSAT program decreases the interval between digits (ISI) by 20 ms whenever a correct response occurs. Each incorrect response results in an ISI increase of 20 ms. This is thought to reduce potential frustration and stress, as the examinee does not have to endure a large number of trials at an interval that is clearly beyond his or her ability. Frustration is also reduced by having all sums range from 2 to 10, rather than the 3 to 16 of the original PASAT. The main performance variables recorded for analyses are the number of correct responses made and the smallest ISI obtained. Information on the ISI during the last trial, number of incorrect responses, and number of missed responses are also readily available.

Victoria Stroop Test

Scores on the Stroop test²⁷ reflect the ability to suppress an automatic reading response and to shift perceptual sets to conform to changing demands while under a time constraint. The Victoria version takes 5 minutes to administer. Participants are shown three cards with 24 dots, words, or color names (six rows of four items) per card. On each card, the stimuli are colored in red, green, blue, and yellow ink in pseudo-random order, and the participant must name the color of each dot, word, or color name, moving from left to right, as quickly as possible. On card C, the color names are printed in a color that is never the same as the word (eg, the word “yellow” printed in blue ink). For each card, the time taken to say the correct color of all items on the card (in seconds) is recorded. Thus, scores provide an indication of processing speed with varying cognitive loads. A discrepancy score is obtained by comparing time taken to finish part C with the baseline dots condition. Test-retest reliability coefficients of 0.90, 0.83, and 0.91 have been reported for the three parts of the Victoria Stroop test, and trial-retrial reliability for this test is above 0.75.²⁸

Procedure

Ethical approval for this study was obtained from the national Health and Disability Ethics Committee. A participant information sheet, a consent form, and a postage-paid envelope addressed to the researcher were mailed to all individuals included in the distribution list of the Auckland branch of the New Zealand Multiple Sclerosis Society. A total of 400 such packages were dis-

tributed, of which 61 (15.25%) were returned. Of those consent forms returned, three were not completed and six were from individuals who provided consent but could not be contacted despite repeated attempts, resulting in a total sample of 52 participants. Although the information was geared toward individuals with MS, the distribution list included not only those with a diagnosis of MS, but also their spouses, clinicians working with MS patients, and people who had made donations to the Society. The Society estimated that about 45% of the distribution list (ie, 180 recipients) would be people who had received a diagnosis of MS. If this figure is accurate, then the return rate becomes 33.9%.

Using telephone numbers requested on the consent form, the researcher contacted participants to schedule individual assessment sessions. Although an accessible assessment room on university premises was available, at the participants' request all assessments were conducted in participants' homes because of identified difficulties with mobility and access to transportation. During each assessment, attempts were made to reduce distractions—for example, by having only the researcher and the participant present, and situating the assessments at either a kitchen or a dining room table such that the participant faced a wall. Each session began with a review of the participant information sheet, including the confidential and voluntary nature of participation, and the ability to withdraw from the study at any time. Prior to cognitive testing, each participant was interviewed to obtain demographic information, information on MS status (eg, when it was diagnosed), and current level of functioning as assessed using the EDSS. Cognitive assessments were then completed including the PASAT, the A-PSAT, and the Victoria Stroop test.

Administration of all tests was in accordance with standardized procedures. Time to complete assessments ranged from 40 to 67 minutes. Follow-up assessments were conducted approximately 6 months after baseline assessments (mean, 6.03; SD = 0.12). Following completion of all sessions, assessments were scored according to standardized procedures.

Statistical Analyses

All data were then entered into an SPSS file (version 17.0; SPSS, Chicago, IL) for analysis. To describe the overall performance of the sample at each measurement time, means and standard deviations of performance (raw scores) were determined for each test administered at baseline and 6-month follow-up, as were the means

and standard deviations of difference between baseline and follow-up scores for each measure. A series of repeated-measures within-subject comparisons with Bonferroni correction for multiple comparisons was then conducted to determine whether change from baseline to 6-month follow-up was significant.

Individual-level change scores were examined in two ways. First, the SD method was used, calculated as $(X_2 - X_1)/SD$, where X_1 and X_2 are the individual's observed baseline and 6-month follow-up scores. Corrections for measurement error and practice effects were then calculated for each participant using the RCI, calculated as $([X_2 - X_1] - [M_2 - M_1])/SD$, where X_1 and X_2 are the individual's observed baseline and 6-month follow-up scores, M_1 and M_2 are the group mean baseline and follow-up scores, and SD is the standard deviation of the group baseline–follow-up difference score.¹⁷ For each of the two methods, reliable change was defined both 1) as a 1 SD change in performance, and 2) with α set at 0.10 (two-tailed; 90% confidence of improvement), which requires a more conservative 1.645 SD change in performance.

Results

Overall Functioning

Baseline EDSS status of the sample ranged from 0 (normal) to 8 (essentially restricted to bed or a wheelchair), with a mean score of 3.5. Those with RRMS had a mean initial EDSS score of 2.50 (SD = 2.53), while those with PPMS had a mean score of 5.23 (SD = 2.52), indicating a greater level of initial disability. Mean EDSS score at 6-month follow-up had increased

to 4.8, indicating a decline in neurologic functioning. The RRMS and PPMS groups had mean follow-up scores of 3.86 (SD = 2.50) and 6.28 (SD = 2.19), respectively. Repeated-measures contrasts indicate that the overall level of within-subject change from baseline to 6 months was statistically significant ($F_{1,45} = 26.802$, $P < .001$), and that the extent of change also differed significantly between the groups ($F_{1,45} = 4.295$, $P = .044$). This between-groups difference supports the analysis of reliable change separately by group.

Table 1 shows means and standard deviations for scores on each neuropsychological test administered at baseline and 6-month follow-up, including the results of within-subject repeated-measures analyses of variance (ANOVAs) used to determine the statistical significance of change over time. As the focus here is on within-subject change, only data for those participants who completed both baseline and follow-up assessments are considered in the analyses; the N for each assessment is shown in the table. Complete data for each measurement point by MS type are presented in Table 2. Only participants tested on both occasions were included in the ANOVA; this is reflected in the degrees of freedom for each ANOVA result in Table 2. Although participants tended to have worse performances at 6-month follow-up, the only changes that were statistically significant based on ANOVA were an increase in the number of correct responses provided for the 2.0-second PASAT trial for participants overall (N = 15; Table 1) and for the RRMS group analyzed separately (N = 19; Table 2). Thus, although no change was seen for the 2.4-second PASAT trial, scores improved significantly for the

Table 1. Raw scores for baseline, 6-month follow-up, and change over time with significance of change over time

Measure	Baseline		6-month follow-up			Difference		Within-subject significance of change	
	Mean	SD	Mean	SD	N	Mean	SD	F	P
PASAT, No. correct									
2.4"	35.06	11.27	31.75	14.01	26	-1.69	10.36	0.693	.413
2.0"	28.13	10.34	35.05	12.80	15	6.67	6.49	15.837	.001
A-PSAT, smallest ISI, ms	1714.78	484.82	1955.0	187.96	42	291.91	1818.06	1.083	.304
Stroop test, s									
Dots	14.99	5.13	15.94	8.82	47	1.09	5.80	1.651	.205
Words	19.73	7.96	18.29	10.01	47	-1.02	6.81	1.047	.312
Colors	32.4	20.18	35.33	33.13	47	2.30	23.93	0.435	.513

Abbreviations: A-PSAT, Simple Adjusting–Paced Serial Addition Test; ISI, interstimulus interval; PASAT, Paced Auditory Serial Addition Test; SD, standard deviation.

Note: For each task, N is the number of participants who completed both baseline and follow-up assessments.

Table 2. Raw scores for baseline, 6-month follow-up, and change over time with significance of change by MS type

Measure	Baseline			6-month follow-up			Difference		Within-subject significance of change	
	Mean	SD	N	Mean	SD	N	Mean	SD	F	P
<i>Relapsing-remitting MS</i>										
PASAT, No. correct										
2.4"	36.05	11.73	19	34.56	13.44	14	-1.71	12.17	0.334	.571
2.0"	28.89	11.74	19	37.93	10.61	24	6.60	5.08	16.869	.003
A-PSAT, smallest ISI, ms	1582.07	464.83	12	1586.20	790.62	29	4.14	551.71	0.002	.968
Stroop test, s										
Dots	14.25	5.56	28	14.31	4.38	29	0.03	4.00	0.003	.960
Words	18.19	6.81	28	17.14	5.85	29	-1.40	3.90	3.616	.068
Colors	29.65	13.92	28	28.62	12.35	29	-1.29	7.94	0.741	.397
<i>Primary progressive MS</i>										
PASAT, No. correct										
2.4"	33.71	10.89	14	27.07	14.12	15	-1.67	6.28	0.633	.449
2.0"	27.00	8.20	13	29.29	16.63	7	6.80	9.45	2.592	.183
A-PSAT, smallest ISI, ms	1941.18	443.37	6	2668.00	2955.08	15	933.85	3152.22	1.141	.306
Stroop test, s										
Dots	15.95	5.74	22	18.42	12.75	19	2.63	7.60	2.281	.148
Words	21.69	9.01	22	20.05	14.24	19	-0.45	9.75	0.040	.843
Colors	35.91	26.04	22	45.58	49.46	19	7.60	36.33	0.832	.374

Abbreviations: A-PSAT, Simple Adjusting-Paced Serial Addition Test; ISI, interstimulus interval; MS, multiple sclerosis; PASAT, Paced Auditory Serial Addition Test; SD, standard deviation.

Note: The difference in N between baseline and 6-month follow-up reflects variations in participants' ability to complete the task at the two time points.

more difficult 2.0-second PASAT task. This result could reflect differences in the group of participants who were able to complete the task on each occasion.

Table 3 presents the means and ranges of change scores obtained using the SD and RCI methods of calculation as well as an indication of the number and proportion of individuals identified as having reliable improvement or decline for each method. The two methods of calculation produced similar levels of reliable improvement for the 2.4-second PASAT and Stroop words trial, while a greater proportion of participants improved reliably when using the SD method for the A-PSAT, Stroop dots, and Stroop colors trials. In all cases, with the exception of the PASAT trials, use of the RCI method identified fewer cases of individual decline than the SD method.

Functioning by MS Type

Table 2 shows means and standard deviations on each test administered at baseline and 6-month follow-up by MS type, including the results of within-subject repeated-measures ANOVAs used to determine the statistical

significance of change over time for those participants who were assessed twice on a task. As the table shows, those with RRMS had better performances at both baseline and 6-month follow-up than those with PPMS. Performance of those with RRMS improved significantly from baseline to follow-up for the 2.0-second PASAT trial. Performance on Stroop trials also improved at follow-up, with the difference approaching statistical significance for the words trials. There was no noticeable change in the A-PSAT. For those with PPMS, performances were worse at follow-up than at baseline on the A-PSAT and on two of the three Stroop trials; however, none of the changes over time were significant.

Table 4 presents the means and ranges of change scores obtained using the SD and RCI methods of calculation as well as an indication of the number and proportion of individuals identified as having reliable improvement or decline for each method and at both ± 1 SD and ± 1.645 SD levels of change. The data indicate that when using the RCI method, few individuals with either RRMS or PPMS declined reliably over the 6-month follow-up period, while use of the SD method

Table 3. Change from baseline to 6 months with proportions improved and declined calculated using SD and RCI methods

Measure	SD method ^a							RCI method ^b					
	Mean	Min, Max	±1 SD		±1.645 SD		Mean	Min, Max	±1 SD		±1.645 SD		
			N, % improve	N, % decline	N, % improve	N, % decline			N, % improve	N, % decline	N, % improve	N, % decline	
PASAT													
2.4"	-0.150	-2.40, 1.51	5, 26.7%	3, 11.5%	0	2, 7.7%	0.156	-2.29, 1.96	4, 15.3%	3, 11.5%	2, 7.7%	2, 7.7%	
2.0"	0.645	-0.68, 1.55	2, 7.7%	0	0	0	-0.039	-2.15, 1.40	2, 13.3%	1, 6.7%	0	1, 6.7%	
A-PSAT ^c	0.602	-1.11, 23.51	3, 7.2%	1, 2.4%	3, 7.2%	0	0.030	-0.43, 6.14	2, 4.2%	0	1, 2.4%	0	
Stroop test													
Dots	0.212	-2.13, 4.88	5, 10.6%	2, 4.2%	4, 8.4%	2, 4.2%	0.025	-2.04, 4.15	6, 12.6%	2, 4.2%	2, 4.2%	1, 2.1%	
Words	-0.128	-1.59, 4.66	1, 2.1%	3, 6.3%	1, 2.1%	0	0.062	-1.65, 5.66	2, 4.2%	2, 4.2%	1, 2.1%	1, 2.1%	
Colors	0.114	-2.43, 6.94	2, 4.2%	3, 6.3%	2, 4.2%	1, 2.1%	-0.026	-2.17, 5.73	2, 4.2%	2, 4.2%	1, 2.1%	1, 2.1%	

Abbreviations: A-PSAT, Simple Adjusting-Paced Serial Addition Test; PASAT, Paced Auditory Serial Addition Test; RCI, Reliable Change Index; SD, standard deviation.

^aCalculated as (follow-up minus baseline)/SD.

^bCalculated as $([X_2 - X_1] - [M_2 - M_1]) / SD_{\text{difference}}$.

^cA-PSAT score used is smallest interstimulus interval.

resulted in reliable declines for a small number of individuals on most measures. In terms of reliable improvements in performance, those with RRMS were less likely to improve reliably than those with PPMS when the RCI method of calculation was used. The proportion of those improving reliably was greater for both MS type subgroups when the SD method was used.

Discussion

The main goal of MS treatment is improvement in the patient's symptoms or at least achieving stability of the disease's impact.²⁹ Yet the measurement of reliable change at the individual level in people with MS remains a significant challenge. This study was conducted to examine stability or change in performance of people with MS on commonly administered tests using the SD and RCI methods with a 6-month test-retest interval. Most reassessments of MS patients receiving interferon beta (IFN β) treatment occur at 3- to 6-month intervals.³⁰

This study found differences in the results of using the two different methods for determining individual change in MS patients. The simple SD method assumes a change in performance if the difference score exceeds the group mean of the difference scores by more than 1 SD. The RCI method is intended to take better account of practice effects and other sources of variance when determining change in performance over time, and typically uses a more conservative 90% confidence level (± 1.645 SD). The RCI method focuses on the

distribution of change scores and includes the measure's test-retest reliability.³¹ One study³² compared six change score methods for examining data from the Canadian Study of Health and Aging, including the SD and RCI methods. They concluded that the RCI is useful in the diagnostic discrimination of developing dementia versus remaining cognitively healthy among older adults.

In our study, the only statistically significant change in group means was an increase in the number of correct responses produced for the 2.0-second PASAT trial. The PASAT has been recommended for use in MS clinical trials.³³ Differences at a group mean level do not address the clinical question of identifying the extent to which people can be expected to reliably improve or decline. Changes were seen at an individual level at the 6-month follow-up. Participants tended to have worse performances at the 6-month follow-up, which is consistent with the decline in neurologic abilities indicated by increased EDSS scores. With the exception of the PASAT trials, use of the RCI method identified fewer cases of reliable decline than the SD method. The 2.0-second PASAT would be considered among the most cognitively challenging tests used. The responsiveness of the PASAT to cognitive decline in the current sample fits with and supports the inclusion of an abbreviated version of the PASAT in the MSFC developed by the National Multiple Sclerosis Society Task Force on Clinical Outcomes.^{12,13} As previously reported,³⁴ the MSFC is more sensitive than the EDSS in demonstrating change. Individuals with MS have consistently been

Table 4. Change from baseline to 6 months with proportions improved and declined calculated using SD and RCI methods by MS type

Measure	SD method ^a							RCI method ^b					
	Mean	Min, Max	±1 SD		±1.645 SD		Mean	Min, Max	±1 SD		±1.645 SD		
			N, % improve	N, % decline	N, % improve	N, % decline			N, % improve	N, % decline	N, % improve	N, % decline	
<i>Relapsing-remitting MS</i>													
PASAT													
2.4"	-0.151	-2.40, 1.51	2, 11.8%	3, 17.6%	0	2, 11.8%	0.155	-2.29, 1.96	2, 11.8%	3, 17.6%	1, 5.9%	2, 11.8%	
2.0"	0.638	0.29, 1.55	2, 20%	0	0	0	-0.049	-0.60, 1.40	2, 20.0%	0	0	0	
A-PSAT ^c	0.009	-1.11, 4.99	2, 6.8%	1, 3.4%	2, 6.8%	0	-0.128	-0.43, 1.20	1, 3.4%	0	0	0	
Stroop test													
Dots	0.008	-2.13, 2.23	1, 3.6%	2, 7.1%	1, 3.6%	2, 7.1%	-0.156	-2.04, 1.81	1, 3.6%	2, 7.1%	1, 3.6%	1, 3.6%	
Words	-0.176	-1.59, 0.80	0	1, 3.4%	0	1, 3.6%	0.005	-1.65, 1.15	1, 3.4%	1, 3.4%	0	0	
Colors	-0.064	-1.05, 0.65	0	2, 7.2%	0	0	-0.177	-1.00, 0.42	0	1, 3.6%	0	0	
<i>Primary progressive MS</i>													
PASAT													
2.4"	-0.148	-0.80, 0.98	0	0	0	0	0.159	-0.55, 1.38	1, 5.3%	0	0	0	
2.0"	0.658	-0.68, 1.45	3, 12.9%	0	0	0	-0.019	-2.15, 1.24	1, 20%	1, 20%	1, 20%	1, 20%	
A-PSAT ^c	1.926	-0.78, 2.51	1, 7.75%	0	1, 7.75%	0	0.383	-0.34, 6.14	1, 7.7%	0	1, 7.7%	0	
Stroop test													
Dots	0.513	-0.90, 4.88	5, 26.3%	0	3, 15.8%	0	0.292	-0.95, 4.15	5, 26.3%	0	2, 10.5%	0	
Words	-0.056	-2.43, 6.94	2, 10.5%	1, 5.3%	1, 5.3%	0	0.145	-1.21, 5.66	1, 5.3%	1, 5.3%	1, 5.3%	0	
Colors	0.377	-1.22, 4.66	2, 10.5%	1, 5.3%	2, 10.5%	1, 5.3%	0.195	-2.17, 5.73	2, 10.5%	1, 5.3%	1, 5.3%	1, 5.3%	

Abbreviations: A-PSAT, Simple Adjusting-Paced Serial Addition Test; MS, multiple sclerosis; PASAT, Paced Auditory Serial Addition Test; RCI, Reliable Change Index; SD, standard deviation.

^aCalculated as (follow-up minus baseline)/SD.

^bCalculated as $([X_2 - X_1] - [M_2 - M_1]) / SD_{\text{difference}}$.

^cA-PSAT score used is smallest interstimulus interval.

found to show cognitive impairment using the PASAT, leading it to be recommended as a core outcome measure in clinical trials.

Only people with RRMS receive IFN β treatment, and changes in their capabilities are typically tracked with the MSFC.³³ The MSFC is used as a primary outcome measure in placebo-controlled trials of IFN β treatments for MS patients.³⁴ The MSFC was designed to be multidimensional, with measures that change independently over time, and to include one measure of change in cognitive functioning. It includes a PASAT trial that initially has a 3.0-second ISI, with the option of a 2.0-second version. One problem with use of the MSFC is the lack of consensus on what constitutes a clinically important change, both on the overall score and on the subscores. It has been previously reported that the MSFC is sensitive to change at intervals of 1 to 2 years.³⁵ The present study showed individual changes in PASAT

performance (using ISIs of 2.4 and 2.0 seconds) over a 6-month period.

When the SD method was used, several participants showed reliable improvement in the 2.0- and 2.4-second PASAT trials, A-PSAT, Stroop dots, and Stroop colors trials. Fewer reliable improvements were seen using the RCI. The two methods of calculation produced similar levels of reliable improvement for the 2.4-second PASAT and Stroop words trials. This suggests greater influence of practice effects on the SD change scores, as it is inconsistent with the overall decline in neurologic ability based on the EDSS scores. The results suggest that the RCI method has addressed the potential influence of practice effects, which would lead to improved rather than reduced performance, but could also indicate that the RCI is too conservative.

Differences in change scores were found between groups. One would expect the PPMS group to decline gradually over time, with the RRMS group show-

ing more variability. None of the RRMS participants were prescribed steroids, which would be expected to improve performance over the course of the study, and none of them suffered a relapse during the course of the study. Overall, both groups had increases in their EDSS scores, indicating greater disability, but the increase was greater in the RRMS group. The SD and RCI methods performed similarly for the PPMS group in identifying individuals showing improvement and decline. The proportion of patients showing reliable improvement on the A-PSAT and all trials of the Stroop test was greater in those with PPMS than in those with RRMS. Given that clinically the PPMS group would be expected to remain stable or decline, it could be hypothesized that improvements in performance seen for some measures in this group reflect the influence of practice effects. The RCI method appears to be effective in accounting for practice effects, as fewer individual improvements were seen with this method. Actual speed of responding in the PPMS group was considerably lower and more variable at 6 months compared with baseline, as measured by the A-PSAT, and thus improvements would not be expected.

Among the limitations of this study was a relatively small sample, which translated to a very small number of individuals showing any change over the 6-month test-retest interval. The sample size was further influenced by the level of cognitive dysfunction of the participants, which reduced the number of individuals able to complete all assessments at each time period. Although none of the participants reported taking medications to improve cognition or experiencing a relapse during the

course of the study, it is possible that some experienced relapse just after participation in the 6-month follow-up assessment; this must be considered in the generalization of the findings.

It is recommended that manufacturers of clinical tests provide information on change over time within clinical samples so that clinicians can calculate the RCI for individual patients. For example, using the RCI method, if an individual with RRMS obtained a score of 35 on the 2.0-second PASAT trial at initial assessment and a score of 42 on the same trial after a 6-month interval, his or her RCI would be calculated as $((42-35)-[37.93-28.89])/11.74 = -0.17$; thus, although the score is improved, the criterion for reliable improvement has not been met. This is particularly important for patients being seen in legal contexts and when seeking funding to support services. Most randomized controlled trials rely on averaging or pooling data across individuals to determine mean change. The responsiveness of individuals to treatment effects can be lost in such analyses, yet such individual responsiveness is important to future efforts to develop and refine treatments. □

Acknowledgments: We would like to acknowledge the New Zealand Multiple Sclerosis Society, Auckland branch, for assistance with this project; and the MS patients who participated in this study, whose time and efforts are much appreciated.

Financial Disclosures: The authors have no conflicts of interest to disclose.

References

1. Portaccio E, Goretti B, Zipoli V, et al. Reliability, practice effects, and change indices for Rao's Brief Repeatable Battery. *Mult Scler*. 2010;16:611-617.
2. Amato MP, Zipoli V, Portaccio E. Cognitive changes in multiple sclerosis. *Exp Rev Neurother*. 2008;8:1585-1596.
3. Rogers JM, Panegyres PK. Cognitive impairment in multiple sclerosis: evidence-based analysis and recommendations. *J Clin Neurosci*. 2007;14:919-927.
4. Brassington JC, Marsh NV. Neuropsychological aspects of multiple sclerosis. *Neuropsychol Rev*. 1998;8:43-77.
5. Bobholz JA, Rao SM. Cognitive dysfunction in multiple sclerosis: a review of recent developments. *Curr Opin Neurol*. 2003;16:283-288.
6. Dineen RA, Vilisaar J, Hlinka J, et al. Disconnection as a mechanism for cognitive dysfunction in multiple sclerosis. *Brain*. 2009;132:239-249.
7. De Sonneville LM, Boringa JB, Reuling IE, Lazeron RH, Ader HJ, Polman CH. Information processing characteristics in subtypes of multiple sclerosis. *Neuropsychologia*. 2002;40:1751-1765.
8. DeLuca J, Chelune GJ, Tulskey DS, Lengenfelder J, Chiaravalloti ND. Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? *J Clin Exp Neuropsychol*. 2004;26:550-562.
9. Demaree HA, DeLuca J, Gaudino EA, Diamond BJ. Information processing speed—a key deficit in multiple sclerosis: implications for rehabilitation. *J Neurol Neurosurg Psychiatry*. 1999;67:661-663.
10. Snyder PJ, Cappelleri JC. Information processing speed deficits may be better correlated with the extent of white matter sclerotic lesions in multiple sclerosis than previously suspected. *Brain Cogn*. 2001;46:279-284.
11. Barak Y, Achirona A. Effect of interferon-beta-1b on cognitive functions in multiple sclerosis. *Eur J Neurol*. 2002;47:11-14.

Practice Points

- Clinicians must be able to determine whether change in an individual's test scores over time represents a significant improvement or decline.
- Two methods for calculating the significance of such change are the Reliable Change Index (RCI) and the standard deviation (SD) method.
- The RCI was found to be a more stringent method, identifying fewer individuals who declined reliably over 6 months.
- The SD method resulted in reliable declines for a small number of individuals on most measures. Use of the SD method resulted in a larger number of individuals who improved reliably.

12. Cutter GC, Baier ML, Rudick RA. Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain*. 1999;122:871-882.
13. Rudick RA, Antel J, Confavreux C. Recommendations from the National Multiple Sclerosis Society Clinical Outcomes Assessment Task Force. *Ann Neurol*. 1997;42:379-382.
14. Barker-Collo SL. Quality of life in multiple sclerosis: does information-processing speed have an independent effect? *Arch Clin Neuropsychol*. 2006;21:167-174.
15. Wise EA. Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *J Pers Assess*. 2004;82:50-59.
16. Bauer S, Lambert MJ, Nielsen SL. Clinical significance methods: a comparison of statistical techniques. *J Pers Assess*. 2004;82:60-70.
17. Troster AI, Woods SP, Morgan EE. Assessing cognitive change in Parkinson's disease: development of practice effect-corrected reliable change indices. *Arch Clin Neuropsychol*. 2007;22:711-718.
18. Barber AP, Hach S, Tippett LJ, Ross L, Merry AF, Milsom P. Cerebral ischemic lesions on diffusion-weighted imaging are associated with neurocognitive decline after cardiac surgery. *Stroke*. 2008;39:1427-1433.
19. Chelune GJ, Naugle RI, Lüders H, Sedlak J, Awad IA. Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology*. 1993;7:41-52.
20. Solari A, Radice D, Manneschi L, Motti L, Montanari E. The multiple sclerosis functional composite: different practice effects in the three test components. *J Neurolog Sci*. 2005;228:71-74.
21. Rosti-Otajarvi E, Hamalainen P, Koivisto K, Hokkanen L. The reliability of the MSFC and its components. *Acta Neurolog Scand*. 2008;117:421-427.
22. Kurtzke JF. Rating neurological impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33:1444-1452.
23. Shawarzyn MA, Schiaffino KM, LaRocca NG, Johnston MV. Determinants of health-related quality of life in multiple sclerosis: the role of illness intrusiveness. *Mult Scler*. 2002;8:310-318.
24. Hobart J, Freeman J, Thompson A. Kurtzke scales revisited: the application of psychometric methods to clinical intuition. *Brain*. 2000;123:1027-1040.
25. Gronwall D. Paced auditory serial-addition task: a measure of recovery from concussion. *Percept Mot Skills*. 1977;44:367-373.
26. Royan J, Tombaugh TN, Rees L, Francise M. The Adjusting-Paced Serial Addition Test (Adjusting-PSAT): thresholds for speed of information processing as a function of stimulus modality and problem complexity. *Arch Clin Neuropsychol*. 2004;19:131-143.
27. Strauss E, Sherman EMS, Spreen O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. New York, NY: Oxford University Press; 2006.
28. Spreen O, Strauss E. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. 2nd ed. New York, NY: Oxford University Press; 1998.
29. D'Souza M, Kappos L, Czaplinski A. Reconsidering clinical outcomes in multiple sclerosis: relapses, impairment, disability and beyond. *J Neurol Sci*. 2008;274:76-79.
30. Limmroth V, Putzki N, Kachuck NJ. The interferon beta therapies for treatment of relapsing-remitting multiple sclerosis: are they equally efficacious? a comparative review of open-label studies evaluating the efficacy, safety, or dosing of different interferon beta formulations alone or in combination. *Ther Adv Neurol Disord*. 2011;4:281-296.
31. Barr W. Neuropsychological testing for assessment of treatment effects: methodologic issues. *CNS Spectrum*. 2002;7:304-306.
32. Frerichs R, Tuokko H. A comparison of methods for measuring cognitive change in older adults. *Arch Clin Neuropsychol*. 2005;20:321-333.
33. Cohen JA, Cutter GR, Fischer JS, et al. Use of the multiple sclerosis functional composite as an outcome measure in a phase 3 clinical trial. *Arch Neurol*. 2001;58:961-967.
34. Cohen JA, Cutter GR, Fischer JS, Goodman AD, Heidenreich FR, Kooijmans MF. Benefit of interferon β -1a on MSFC progression in secondary progressive MS. *Neurology*. 2002;59:679-687.
35. Fischer J, Rudick R, Cutter G, Reingold S. The Multiple Sclerosis Functional Composite Measure (MSFC): an integrated approach to MS clinical outcome assessment. *Mult Scler*. 1999;5:244-250.

KEEP YOUR COOL with Kool Max®

The best-selling, most effective cooling vest for MS... worldwide!




- **LIGHTWEIGHT**
You control the weight from 2 lbs. to 5.5 lbs. by varying the number of packs used.
- **COMFORTABLE AND DISCREET**
Fully adjustable for a custom fit.
- **EFFICIENT AND LONG-LASTING**
Will cool for up to 4 hours and works well in any climate.
- **OPTIMAL COOLING RELIEF***
Packs are placed strategically along the spine. *Patent Pending.
- **THE MOST COST-EFFECTIVE CHOICE**
Affordable vests and accessories. Save money with a cooling kit!

Polar Products is a proud supplier for MS non-profit cooling programs nationwide.



POLAR Products

A leading worldwide manufacturer of body cooling systems since 1984.

Give us a call at 1.800.763.8423 to request a catalog!

To see the most complete line of cooling vests and accessories, visit www.polarproducts.com.