The feasibility, reliability and concurrent validity of the MSReactor computerized cognitive screening tool in multiple sclerosis

Daniel Merlo, David Darby, Tomas Kalincik, Helmut Butzkueven† and Anneke van der Walt†

Ther Adv Neurol Disord

2019, Vol. 12: 1–15

DOI: 10.1177/1756286419859183

Correspondence to:
Daniel Merlo
Department of Neuroscience, Central Clinical School, Monash University, Level 6, 99 Commercial Road, Melbourne, Victoria 3004, Australia
daniel.merlo@monash.edu

David Darby
Eastern Clinical Research Unit, Monash University, Melbourne, Australia

Department of Neurology, Box Hill Hospital, Melbourne, Australia

Florey Institute of Neuroscience and Mental Health, Melbourne, Australia

Tomas Kalincik
Department of Medicine, University of Melbourne, Melbourne, Australia

Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia

Helmut Butzkueven
Department of Neuroscience, Central Clinical School, Monash University, Melbourne, Australia

Eastern Clinical Research Unit, Monash University, Melbourne, Australia

Department of Neurology, Alfred Hospital, Melbourne, Australia

Department of Neurology, Box Hill Hospital, Melbourne, Australia

Anneke van der Walt
Department of Neuroscience, Central Clinical School, Monash University, Melbourne, Australia

Melbourne Brain Centre at RMH, University of Melbourne, Melbourne, Australia

Department of Neurology, Alfred Hospital, Melbourne, Australia

Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia

†These authors share senior leadership.

Abstract

Background: Multiple sclerosis (MS) cognitive tests are resource intensive and limited by practice effects that prevent frequent retesting. Brief, reliable and valid monitoring tools are urgently needed to detect subtle, subclinical cognitive changes in people with MS. Cognitive monitoring over time could contribute to a new definition of disease progression, supplementing routine clinical monitoring.

Methods: MSReactor is a web-based battery that measures psychomotor (processing) speed, visual attention and working memory, using simple reaction time tasks. Clinic-based tasks were completed at baseline and 6 monthly with home testing 1–3 monthly. Acceptability, quality of life, depression and anxiety surveys were completed. We studied its correlation with the Symbol Digit Modalities Test, practice effects, test–retest reliability and the discriminative ability of MSReactor.

Results: A total of 450 people with MS were recruited over 18 months, with 81% opting to complete home-based testing. Most participants (96%) would be happy (or neutral) to repeat the tasks again and just four reported the tasks made them 'very anxious'. Persistence of home testing was high and practice effects stabilized within three tests. MSReactor tasks correlated with Symbol Digit Modalities Test scores and participants with MS performed slower than healthy controls.

Conclusion: MSReactor is a scalable and reliable cognitive screening tool that can be used in the clinic and remotely. MSReactor task performance correlated with another highly validated cognitive test, was sensitive to MS and baseline predictors of cognitive performance were identified.

Keywords: attention, cognition, multiple sclerosis, neuropsychology, processing speed, working memory

Received: 19 December 2018; revised manuscript accepted: 31 May 2019

Introduction

Cognitive impairment affects 40–65% of people with multiple sclerosis (MS), leading to lower rates of employment, social isolation and affected activities of daily living.[1] Cognitive impairment occurs throughout the MS disease course,[2] most commonly impacting information-processing speed, attention, working memory and executive function.[3] In its early stages, cognitive change is, however, difficult to detect, both by clinicians[4] and by standard neuropsychological tests[5] because individuals with cognitive decline will remain within the normal range of standard tests at this time. Complex cognitive batteries and even simpler, adapted tests such as the Brief International Cognitive Assessment for MS[6] require dedicated resources to administer and score, making it impractical to use in under-resourced outpatient clinics. The Symbol Digit Modalities Test (SDMT) is recommended for use as a brief and

valid cognitive screening measure where time is limited.[7] Despite the availability of alternate versions of this test, learning effects still occur and this limits their use in situations where frequent and repeated cognitive screening is required, for example, when monitoring for early signs of a treatment response.[8] Other commonly used cognitive screening tools also lack sensitivity to preclinical cognitive change in MS.[9] This is important, as intervention with disease modifying treatments have the greatest impact on physical disability accumulation if used early in the disease course.[10] The same beneficial effects potentially apply to cognition,[11,12] but conclusive evidence regarding long-term effects of current therapies on cognition is lacking.[13] The ability to perform regular cognitive monitoring in the outpatient clinic is currently an unmet need in MS[14] and requires the development of a screening test that can be repeated frequently, with minimal learning effects. An ideal screening test needs to be brief, interesting and self-administered, in addition to being valid, reliable and sensitive to subtle cognitive changes. Computerized screening tests have the potential to address many of these issues.

Computerized cognitive batteries have gained traction in other fields of neurology[15] and efficiently screen broad cognitive functions such as information-processing speed, attention and working memory.[16] Where early computerized cognitive tests aimed to replicate existing 'pen and paper' tests, recent studies have investigated the basic speed of a response, a measure of information-processing speed. This is a key foundational cognitive domain that can be responsible for impairments in higher cognitive abilities, including working memory and executive function.[17] Computerized cognitive batteries are highly useable,[18] stable and reliable across a range of ages in healthy and impaired populations,[19] can be self-administered and have a relative lack of practice effects due to the ability to generate many alternate versions. In our previous work investigating the use of a computerized battery in MS, the detection (Simple Reaction Time, SRT), identification (Choice Reaction Time, ChRT) and One-Back (OBK) tasks of the CogState brief battery were able to discriminate between 70 MS and 37 healthy controls, with the detection and identification tasks more sensitive to cognitive change over 12 months than the Paced Auditory Serial Addition Test (PASAT).[20] 'MSReactor', adapted from tests made available by uBrain (http://ubrain.com.br), is a web-based battery to monitor cognitive abilities in three commonly affected cognitive domains. In this study, we explored the usability, test–retest reliability and practice effects of the MSReactor battery. In addition, we determined the correlation with the SDMT score and compared performance on the cognitive tasks between MS patients and healthy controls (HCs).

Materials and methods

Participants and recruitment

Adult MS participants were recruited between March 2016 and September 2017 from two tertiary MS clinics in Melbourne, Australia. Inclusion criteria included: (a) diagnosis of relapsing–remitting or secondary-progressive MS; (b) no upper limb, visual, or cognitive deficits that preclude performance on a touch-screen device in the clinic; and (c) willingness to use their own computer or tablet device with internet access for home-based testing. HC participants were recruited *via* community notices, self-enrolled and completed testing *via* the testing website. The study was approved by the relevant Ethics Committees and all participants provided written informed consent.

Study design

A prospective convenience sample of MS participants were enrolled during their outpatient visit and provided with a unique password to access the testing website. Clinic-based testing was completed at baseline and each subsequent clinic visit (approximately 6 monthly). Optional home-based testing was offered to all participants and performed 1–3 monthly. HC participants completed home-based testing only. All participants completed at least one (maximum of two) brief practice test prior to their baseline test and were encouraged to perform a practice test prior to the home-based test. Immediately following completion of the tasks, electronic surveys assessing acceptability, quality of life (QoL), anxiety and depression were presented. Total clinic-based testing time was 12–15 min. Surveys were omitted from home tests, resulting in a testing time of about 5 min. Persistence was encouraged by two automated email reminders (sent 1 week apart) if no scored test (clinic or home) had been recorded for 3 months.

Computerized cognitive battery (MSReactor)

MSReactor is accessible *via* any modern internet browser. The battery consisted of three tasks using a set of universal, very simple stimuli presented in a visual game-like interface, including a psychomotor (processing) speed (SRT) test, a visual attention (ChRT) test and a working memory (OBK) test where participants reacted to soccer balls or custom playing cards appearing on the screen. Participants were required to become familiar with the 'yes' and 'no' buttons and each task displayed a textual instruction screen. For the SRT task, participants pressed the 'Yes' button when they detected a yellow ball appear on the screen. For the ChRT task, participants indicated 'yes' if the ball was red and 'no' if the ball was not red. For the OBK task, participants responded 'yes' if the face-up card was identical to the immediately previous card and 'no' if the card was different to the previous card. The cards presented in the OBK task consisted of combinations of four colours, four shapes and eight numbers, allowing for 128 unique possibilities in stimuli. All tasks had a prestimulus interval of 1000 ms, a 100–5000 ms stimulus presentation followed by a randomly variable poststimulus interval of between 0 and 1000 ms. These measures ensured alternate forms of the tasks were generated. On completion of the tasks, results were uploaded to a central database, automatically analysed, collated with prior results for the same participant and made available for review by the participants treating physician.

Acceptability, quality of life, depression and anxiety surveys

Participants completed an acceptability questionnaire to assess the enjoyability, level of anxiety, engagement, duration and repeatability of the tasks [Supplementary File (a)]. Depression was assessed using the Patient Health Questionnaire (PHQ-9);[21] anxiety using the Penn State Worry Questionnaire (PSWQ);[22] and QoL assessed using the Multiple Sclerosis Quality-of-Life score.[23]

Concurrent validity and discriminative ability

A convenience subset of MS participants without relapse or steroid treatment completed the pen-and-paper version of the SDMT in addition to the MSReactor in the same testing session. To determine the ability of the MSReactor tasks to discriminate between MS patients and controls without MS, the baseline task performance of this subset of participants was compared with the baseline task performance of HC participants and controlled for education attainment.

Data analysis

Descriptive data are presented as mean and standard deviation (SD), median and interquartile range (IQR) where appropriate and frequency data as proportions. Acceptability was recorded on Likert scales, ranging from a negative response (0) to a positive response (10) and recoded to 5-point ordinal dummy variables for analysis. For each task, the speed of performance was the average reaction time (ms) for the first 30 correct responses. Individual performance speeds were log-transformed and mean reaction times calculated. Accuracy was defined as the proportion of correct responses made for each task, normalized with an arcsine square-root transformation.

The probability of discontinuing home testing was assessed using a Cox proportional hazards model, with covariates of age and quartiles of baseline task performance. Correlation between baseline task performance and QoL, depression and anxiety were assessed using a Spearman rank coefficient. To assess baseline associations between task performance and disease and demographic factors, multivariable linear regression was performed with task performance as the dependent variable and age, Extended Disability Status Scale (EDSS) and disease duration as independent variables. The effect of time between repeat testing and the number of completed tests on practice effects was assessed in separate linear mixed-effects models and then together using a multivariate analysis with task performance as the dependent variable. Test–retest reliability was assessed by calculating the concordance correlation coefficient (CCC) between each consecutive pair of tests. To visualize the mean distribution of reaction time over the first 10 repeat tests, a curve was interpolated through each timepoint using nonparametric bootstrap for 10,000 resamples and bias-adjusted confidence intervals calculated from the bootstrapped distributions. The mean first derivative, or slope of a line tangent to the interpolated curve, was calculated for each timepoint and bias-adjusted confidence intervals calculated. One-sample t test was used to compare the first derivative at each timepoint ($n = 10{,}000$) to a hypothesized first derivative mean of zero

Table 1. MS participant characteristics.

	Participants	Withdrawers
	n (%)	*n* (%)
Total	450	17
RRMS	435 (97)	17 (100)
SPMS	15 (3%)	0
Female	338 (75)	12 (70.5)
Age (SD)	43.1 years (11.09)	44.7 years (9)
EDSS; median (IQR)	2 (1–3.5)	2 (1–4)
Disease duration (SD)	13.52 years (8.14)	14.07 (7.56)
Opted to complete home testing	364 (81%)	
Repeated testing within 3 months of enrollment	289 (80%)	
Withdrawn from study	17 (3.8%)	

EDSS, Extended Disability Status Scale; IQR, interquartile range; MS, multiple sclerosis; SD, standard deviation; SPMS, secondary progressive multiple sclerosis; RRMS, relapsing–remitting multiple sclerosis.

($mu = 0$). Performance at the second clinic test (approximately 6 months from baseline) was compared with the preceding home test using a linear mixed-effects model. Devices used to perform home tests were summarized. A general linear model was used to compare baseline performance between MS and controls, with all models controlled for years of education. Raw correlations between MSReactor and SDMT scores were calculated using Pearson's correlation coefficient. Disattenuated correlation coefficients between the latent test scores were then calculated by adjusting for reliability of the MSReactor (following stabilization of learning effect) and previously published reliability data for the SDMT[24] for the equivalent testing epoch.

Results

Participant characteristics

Characteristics of the 450 MS participants who completed baseline clinic tests are shown in Table 1. Of these, 364 (81%) opted to complete additional home testing, with most of these participants (80%) completing a home test within 3 months of baseline. Most participants completing home testing used the Windows operating system (42%), followed by iOS (38%), Macintosh operating system (13%) and 'Other' platform (7%). Seventeen participants (3.8%) withdrew from the study. A subset of 30 MS participants completed the MSReactor tasks and SDMT in the same testing session and the baseline task performance of this subset was compared with the baseline performance of HC participants ($n = 30$).

Home-testing persistence

Home-based testing was discontinued by 40 participants (11%) who reverted to clinic-only testing. In multivariate survival analysis, lower quartile (or slower reaction time) performance on all tasks [SRT: hazard ratio (HR) 1.48; 95% confidence interval (CI) 1.10–1.99; ChRT: HR 1.44; 95% CI 1.08–1.93; and OBK: HR 1.35; 95% CI 1.01–1.80] was significantly associated with greater rates of home-testing discontinuation [Figure 1(a–c)]. In addition, older participants were more likely to persist with home testing.

Acceptability

Acceptability surveys were completed by 438 (97.3%) participants at baseline. Participant-rated acceptability of the cognitive tasks was high and is summarized in Table 2.

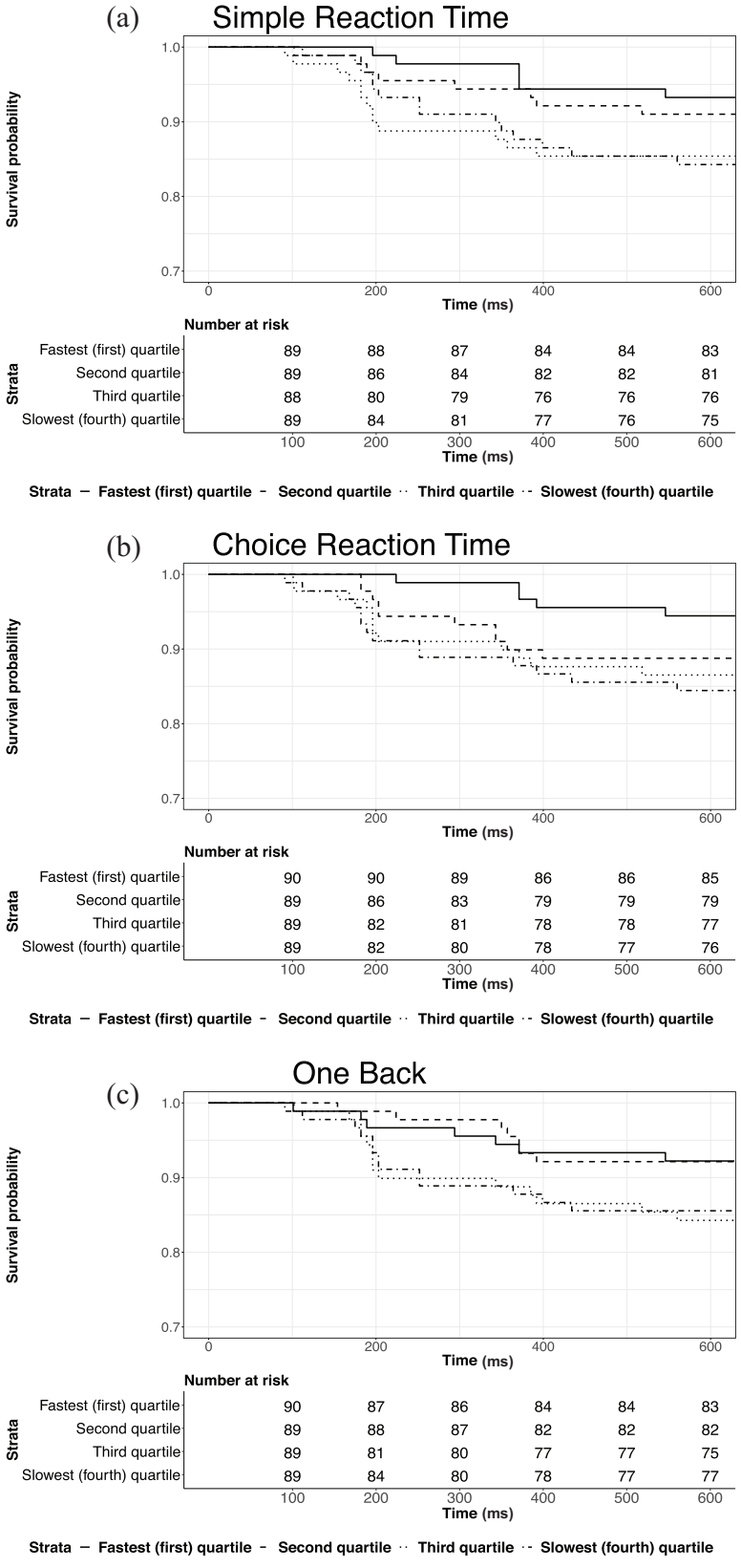

Figure 1. Probability of home-testing persistence based on quartiles of baseline task performance.
(a) Home-testing persistence based on Simple Reaction Time task performance; (b) home-testing persistence based on Choice Reaction Time task performance; and (c) home-testing persistence based on One-Back task performance.

Table 2. Baseline acceptability of the MSReactor tasks.

Did the test make you anxious?	227 (52%)	63 (14.5%)	120 (27%)	24 (5.5%)	4 (1%)	438
	Very much	**A little bit**	**Neutral**	**Not really**	**Not at all**	**Total**
Did you enjoy the test?	79 (18%)	126 (28.8%)	222 (51%)	10 (2%)	1 (0.2%)	438
	Very interesting	**A little bit interesting**	**Neutral**	**Not that interesting**	**Very boring**	**Total**
Did you find the test interesting?	22 (5%)	39 (9%)	317 (72%)	48 (11%)	12 (3%)	438
	Very happy	**Happy**	**Neutral**	**Unhappy**	**Very unhappy**	**Total**
Would you be happy to repeat the test?	197 (45%)	111 (25%)	116 (26%)	7 (2%)	7 (2%)	438
	Too short	**Slightly too short**	**About right**	**Slightly too long**	**Too long**	**Total**
What did you think about the duration of the test?	3 (0.5%)	15 (3.5%)	409 (93%)	7 (2%)	4 (1%)	438

Quality of life, depression and anxiety

Most participants completed baseline QoL (95.5%), depression (94.9%) and anxiety surveys (94.9%). QoL scores correlated weakly with reaction time on the SRT ($r = -0.26$, $p < 0.001$), ChRT ($r = -0.29$, $p < 0.001$) and OBK ($r = -0.26$, $p < 0.001$). PHQ-9 scores correlated weakly with reaction time on the SRT ($r = 0.24$, $p < 0.001$), ChRT ($r = 0.26$, $p < 0.001$) and OBK ($r = 0.26$, $p < 0.001$). PSWQ scores did not significantly correlate with performance on any of the speed measures ($p > 0.05$).

Cognitive performance and baseline predictors

Baseline task performance was independently associated with EDSS and age, but not disease duration (Table 3). For the SRT, ChRT and OBK tasks, each one step increase in EDSS resulted in slowing of the transformed reaction times by between 0.015 and 0.02 log milliseconds, translating to a prolonging of between 13 ms and 25 ms in reaction time per step of increase in EDSS. For each year increase in age, reaction times slowed between 0.001 and 0.002 log milliseconds (or 1 ms and 3.2 ms). Sex was associated with faster reaction times on the OBK task only, with males performing 0.029 log milliseconds (or approximately 44 ms) faster than females.

Learning effects and test–retest reliability

To assess learning effects and test–retest reliability, task performance was examined in MS participants ($n = 328$) who had performed up to 10 successful testing sessions. In this home-testing cohort, the median time interval between tests was 82 days between the first and second test, reducing to 31 days between the second and third test, 29 between the third and fourth test and then stabilizing around 27 days between subsequent tests. In the nonparametric bootstrap fitted data, mean reaction time performance on all tests improved after baseline as evidenced by the slope of the curve being significantly different to the hypothesized mean of zero at baseline ($p < 0.001$). The slope of the fitted curve stabilized rapidly and no more learning effect was evident from the second test for the SRT task and from the third test for the ChRT and OBK tasks, respectively (Figure 2; one-sample t test shown in Appendix 1). The reliability of the tasks improved over time following stabilization of learning effect and the CCC for test 4–5 was 0.77, 0.71 and 0.83; and for tests 8–9 was 0.83, 0.81 and 0.86 for the SRT, ChRT and OBK, respectively (Figure 3; all CCCs shown in Appendix 2). Mean reaction time performance on all tasks at the second clinic testing session was not significantly different from the preceding home test ($p > 0.05$).

Table 3. Multivariable linear regression estimates of the association between baseline patient characteristics and the performance on the MSReactor tasks.

MSReactor task	Independent variable	ß	95% confidence interval	*p* value
Simple Reaction Time	Intercept	2.4963151		
	EDSS	0.018	0.013–0.024	<0.0001*
	Age	0.0014	0.0004–0.0024	0.006*
	Sex (male)	–0.138	–0.14 to 0.01	0.51
	Disease duration	0.0005	–0.0008 to 0.002	0.46
Choice Reaction Time	Intercept	2.6872189		
	EDSS	0.017	0.012–0.022	<0.0001*
	Age	0.001	0.0002–0.002	0.018*
	Sex (male)	–0.01	–0.03 to 0.005	0.16
	Disease duration	0.0003	–0.0008 to 0.001	0.54
One Back	Intercept	2.8354256		
	EDSS	0.016	0.01–0.02	<0.0001*
	Age	0.002	0.0008–0.003	<0.001*
	Sex (male)	–0.029	–0.05 to –0.009	0.005*
	Disease duration	0.0003	–0.0009 to 0.0016	0.61

*Denotes statistical significance.
EDSS, Extended Disability Status Scale.

Concurrent validity and discriminative ability

SDMT scores correlated moderately with SRT performance (Pearson's $r = -0.51$, $p = 0.004$), ChRT performance ($r = -0.59$, $p < 0.001$) and OBK performance ($r = -0.43$, $p = 0.015$; Figure 4). Disattenuated correlation coefficients were $r_{dis} = -0.68$, $r_{dis} = -0.73$ and $r_{dis} = -0.50$, respectively.

MS ($n = 30$) and HC participants ($n = 30$) were well balanced with regards to age [MS mean 41.5 years (SD 11.13) and HC mean 38 years (SD 14.25)], sex [77% (23/30) female and 72% (13/18) female] and years of education [MS mean 15 years (SD 2.72) and HC mean 16.4 years (SD 2.53)] respectively. The mean baseline difference between MS and HC participants for the SRT, ChRT, and OBK tasks was −59.5 ms (95% CI 28–94 ms, $p < 0.001$), −89 ms (95% CI 23–162 ms, $p = 0.01$) and −127 ms (95% CI 21–249 ms, $p = 0.02$), respectively, independent of years of education.

Discussion

To our knowledge, this study is the first to investigate the feasibility of implementing a web-based computerized cognitive screening tool in both the clinic-based and home-based setting for MS. We studied the usability (acceptability, efficiency, stability[18]) and concurrent validity of a computerized cognitive screening platform, MSReactor. Assessing the usability of the battery is an important first step in defining its utility in the clinic setting. Any test that uses an individual's previous test scores to detect subtle change in cognition needs to be administered regularly. Factors that maintain a patient's motivation for testing are therefore critical and the task needs to be brief, nonanxiety provoking and reasonably interesting to perform. Participant response to MSReactor tasks were favourable, with most being happy to repeat the testing and the majority indicating that they thought the duration of the tasks was 'about right'. Only a small fraction of participants found that the tasks made them feel anxious, in contrast to prior studies with tests such as the PASAT, which is frequently reported as aversive and stressful.[25]

Implementation of MSReactor is uncomplicated and allows rapid recruitment of large groups of participants. In this study, it allowed 450 participants to be enrolled by a single, nonexpert

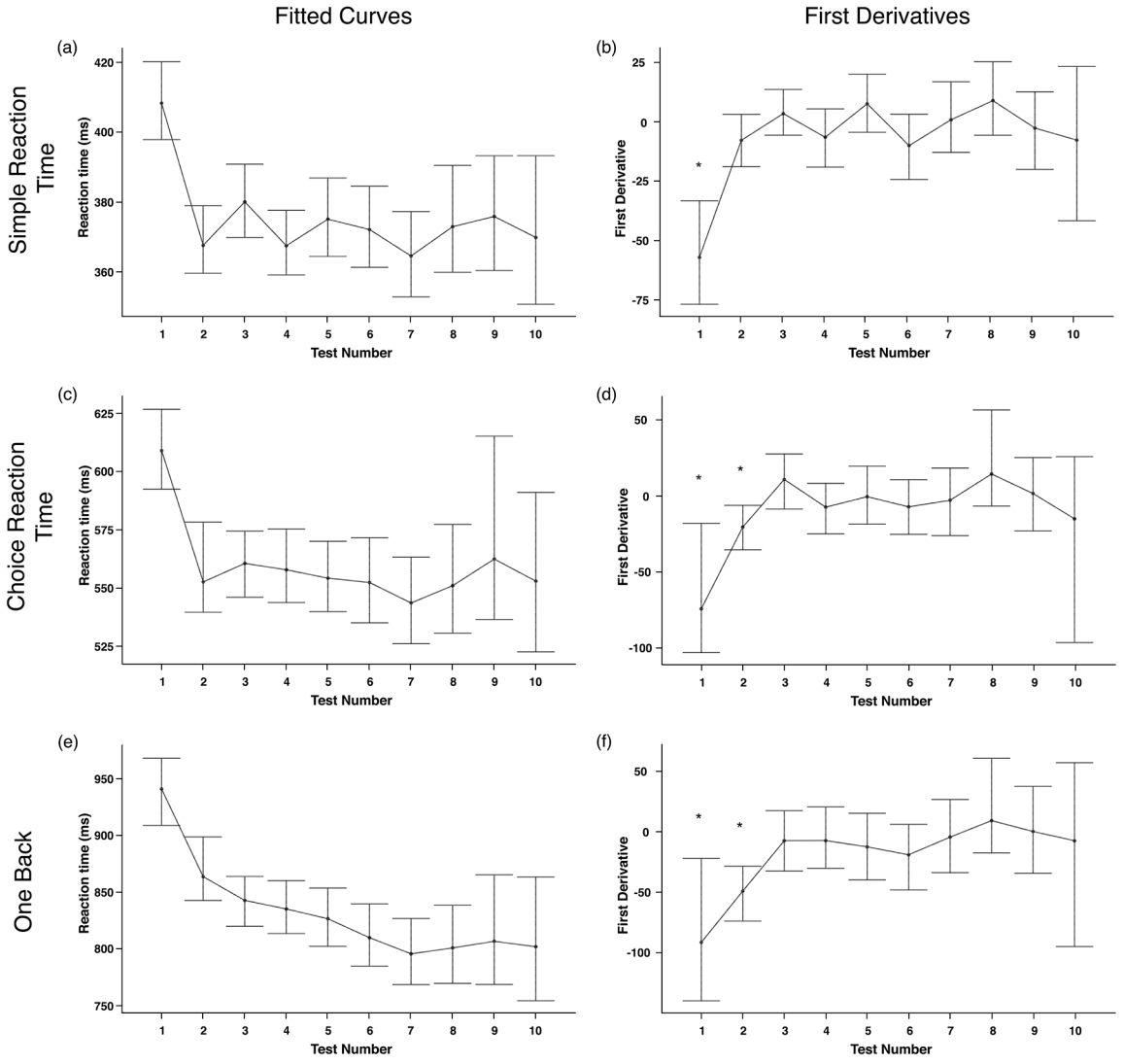

Figure 2. Fitted curves and first derivatives for each of Simple Reaction Time, Choice Reaction Time and One-Back memory task reaction time.

Cubic splines were fitted to the distribution of the first 10 tests for each task using nonparametric bootstrap and bias-corrected confidence intervals calculated (a, c, e). The mean first derivative was calculated for each timepoint and bias-adjusted confidence intervals calculated for each timepoint (b, d, f).

*Indicates timepoints where H0 is rejected ($p < 0.05$) in one-sample t test ($\mu = 0$).

member of the research team over 18 months at just two clinic sessions per week. The brief testing time of 5–15 min and self-administration of the battery means most participants were able to complete the testing, on their own, while waiting for their clinical consultation with no extra time required. This ease of use and lack of requiring a technical support person[26] is a major practical advantage that makes MSReactor suitable for use in a busy tertiary MS clinic.

The majority of participants chose to enrol and also persisted with home testing over time. Benefits of home testing include testing in a familiar or remote environment and allowing frequent testing. This can increase fidelity of serial assessments and should enable earlier detection of change. Home-testing performance over time was equivalent to repeat outpatient clinic testing. The ability to complete testing on a range of everyday electronic screen devices reduced the barrier to remote testing and did not affect the overall performance measures. On the other hand, disadvantages of home testing could include testing in a variable environment, technical support challenges and the possibility of tester substitution.

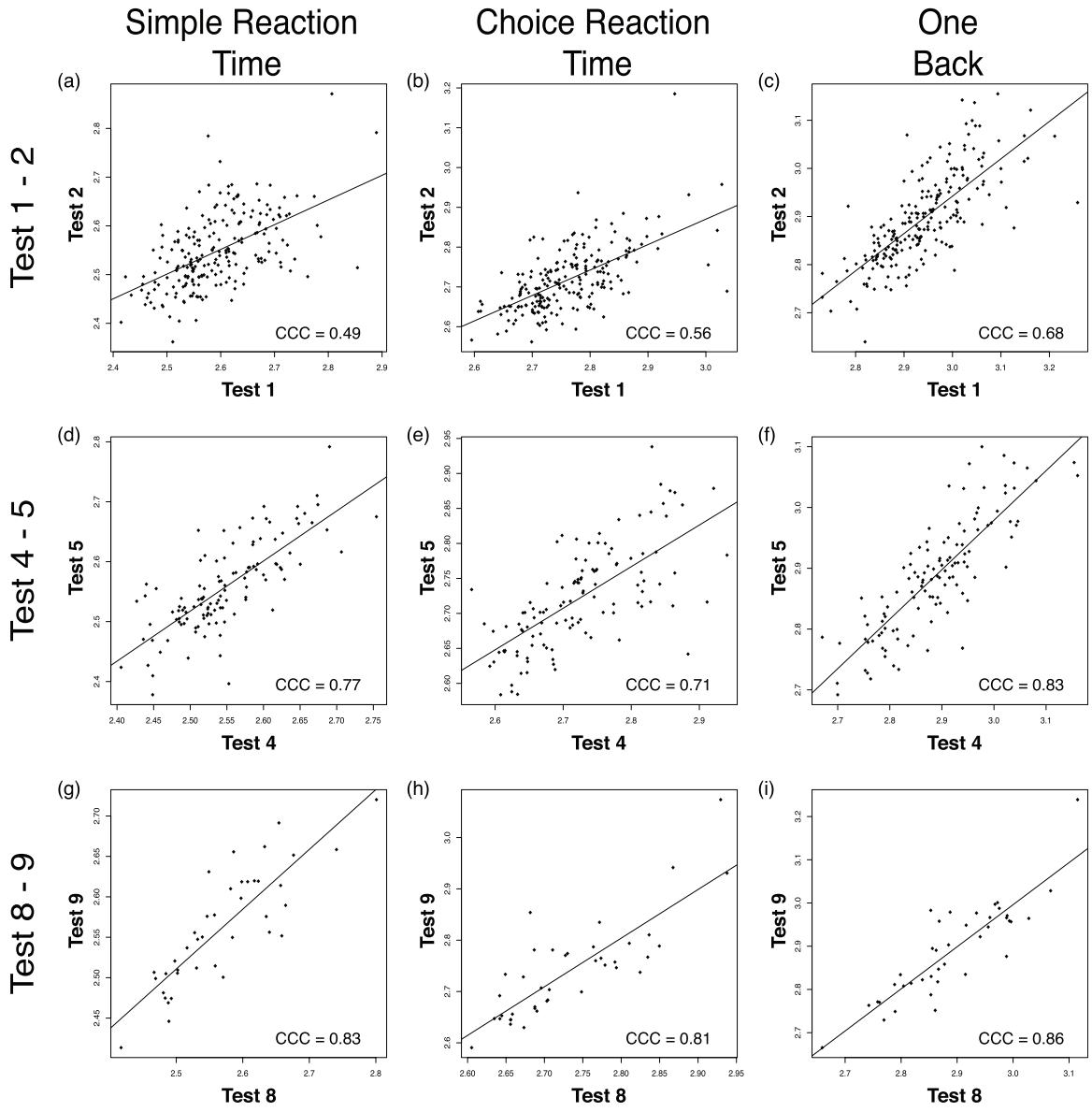

Figure 3. Test–retest reliability.
The concordance correlation coefficient (CCC) was calculated for performance between consecutive pairs of tests for the Simple Reaction Time, Choice Reaction Time and One-Back memory tasks. The CCC improves over time from between test 1 and 2 (a, b, c); to tests 4 and 5 (d, e, f) and tests 8 and 9 (g, h, i).

Although compliance for home testing was high during the follow-up period, 40 participants (11%) chose to revert to clinic-only testing. Interestingly, younger participants were less likely to persist with home testing than older participants, a difference possibly attributable to age-related lifestyle and social differences. Poorer baseline performance on MSReactor tasks was also associated with lower home-testing persistence and possibly reflects lack of motivation, frustration, or apathy.[27] Identification of patients who are noncompliant with remote testing could prompt more detailed cognitive evaluation, in addition to offering tailored support to improve testing persistence, including increased email reminders or mobile-phone-optimized platforms.

Practice effects can be evident in cognitive measurement tools where regular use leads to improvements in test scores in the absence of neurological change. Although practice effects were not eliminated completely with the

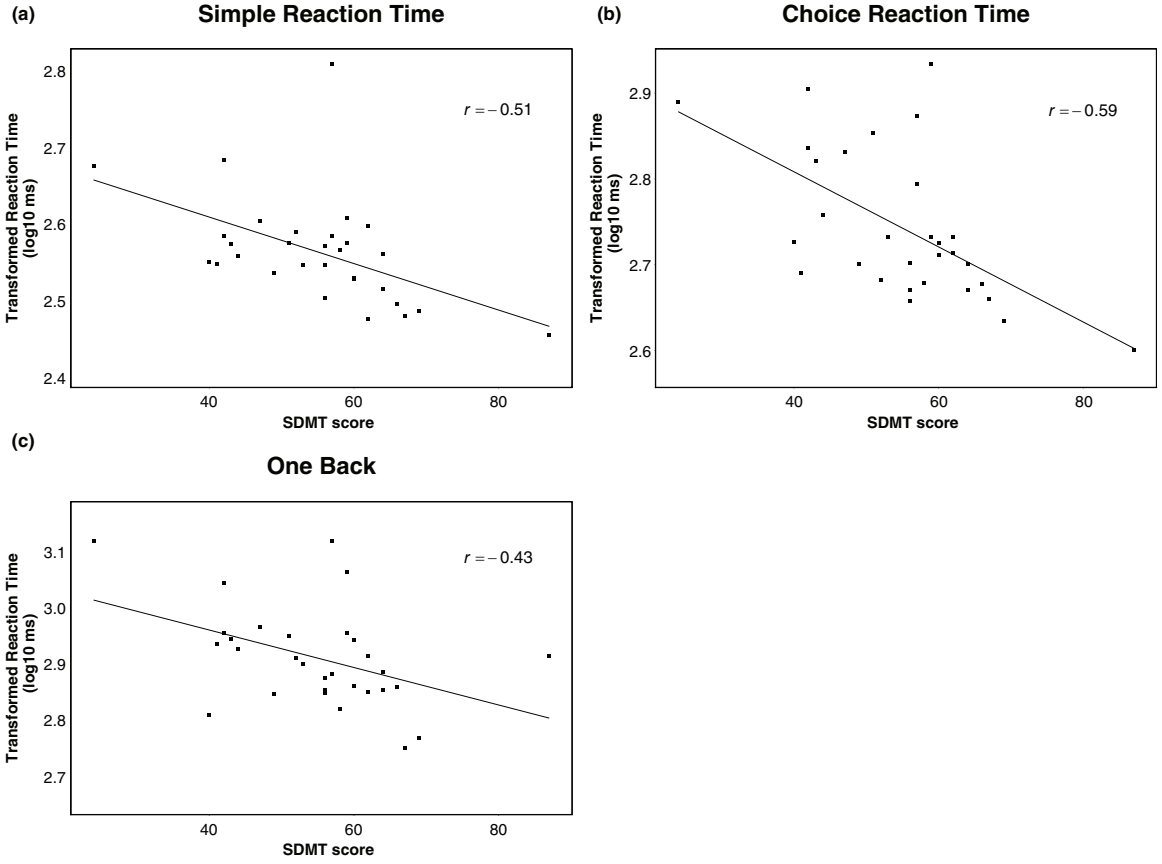

Figure 4. Correlations between SDMT and MSReactor tasks.
Pearson product-moment correlation coefficient was calculated for a subset of participants ($n = 30$) who completed the MSReactor battery and the Symbol Digit Modalities Test in the same testing session. Pearson's r for the Simple Reaction Time (a), Choice Reaction Time (b) and One-Back (c) tasks are shown.
SDMT, Symbol Digit Modalities Test.

MSReactor computerized battery, the learning curve is steep and task performance stabilized within two to three retests, with subsequent high test–retest reliability demonstrated. Task performance correlated only weakly with depression and QoL scores, but not with anxiety. The ability to perform regular testing to identify and quantify the practice effects using a computerized battery is an advantage to standard tools where limited number of alternate versions restrict retest frequency. In a recent study of a computerized version of the SDMT, the Processing Speed Test (PST), Rao and colleagues found significant practice effects in both MS patients and HCs when administered across two sessions (2–3 h apart); however, the persistence of these practice effects in subsequent testing was not explored.[28] Like the PST, the MSReactor tasks demonstrated excellent test–retest reliability following the second administration of the tasks, coinciding with a shorter intertest interval.

MSReactor task performance and SDMT scores were moderately correlated. The SDMT is a commonly used, valid and reliable tool that correlates with lesion burden and brain atrophy,[29,30] yet despite these advantages, the SDMT remains impractical to administer in a busy outpatient clinic. Self-administered computerized cognitive batteries such as MSReactor and the PST may be able to address this limitation. The CogState brief battery, a computerized battery employing a similar testing paradigm to MSReactor, was shown to be construct valid, with the strongest associations between the identification task (processing speed) and the SDMT.[16] Although the MSReactor cognitive tasks described here do not interrogate just a single neuropsychological

construct [psychomotor (processing) speed, visual attention], the good concurrent correlations with the SDMT provide preliminary evidence of measuring comparative neuropsychological functions. Further work is planned to comprehensively validate the MSReactor battery.

The MSReactor tasks were able to discriminate between MS participants and those without MS. Performance on any cognitive task can be influenced by demographics such as educational attainment, age and sex; thus, any meaningful interpretation of cognitive impairment from a test battery must be derived from standardized scores based on normative values. Although the ultimate aim of a screening tool such as MSReactor is to monitor for cognitive change within an individual, where demographics do not change, collection of normative data from people without MS remains a focus of current work.

This study had some limitations. Participation in the study was limited to (predominantly) participants with relapsing–remitting MS (RRMS). We are now broadening the population to include clinically isolated syndrome (CIS), as cognitive impairment is present in up to 30% of patients with CIS. As early intervention with disease-modifying therapies has the greatest impact on disability trajectories, we predict that detection of cognitive change in periods of pretreatment observation or during early therapy in CIS and early RRMS is most likely to improve long-term outcome.

MSReactor is an innovative and self-administered web-based cognitive battery which is highly scalable, well accepted and reliable, suggesting it should be evaluated further as a cognitive screening tool in MS. It is important to note that computerized cognitive batteries are not intended to replace neuropsychological testing but to act as sensitive screening tools that can prompt further clinical testing.[31] Having a brief self-administered monitoring tool could also provide the treating team and the patient with an earlier indication of subtle changes or cognitive relapses. If confirmed using neuropsychological testing, this could lead to early intervention with education on coping strategies and positive efforts to maintain employability. The results from this study forms the basis of future research to define cognitive trajectories across the MS disease course and impact of treatment change on these trajectories.

Acknowledgements

DM gratefully acknowledges the support provided by Australian National Health and Medical Research Council (NHMRC) Research Training Programme scholarship and the support provided by Multiple Sclerosis Research Australia Postgraduate Scholarship. AvW is funded by NHMRC Fellowship 1123330. The investigators gratefully acknowledge: the support provided by Novartis Australia; uBrain for providing the tests; and all of the participants who generously donated their time. HB and AvW are the senior authors of this work.

Funding

The authors disclosed receipt of the following financial support for the research, authorship and/or publication of this article: this work was supported by the Novartis Australia (grant number CFTY720DAU08T).

Conflict of interest statement

DM has received travel support and speaker honoraria from Novartis.

DD is consultant to UBrain, former founder and shareholder of CogState, CEO of Cerescape and received honoraria for lectures from Biogen, Novartis and other pharma.

TK served on scientific advisory boards for Roche, Genzyme-Sanofi, Novartis, Merck and Biogen, steering committee for Brain Atrophy Initiative by Genzyme, received conference travel support and/or speaker honoraria from WebMD Global, Novartis, Biogen, Genzyme-Sanofi, Teva, BioCSL and Merck and received research support from Biogen.

HB's institution (Monash university) has received compensation for consulting, talks, advisory / steering board activities from Biogen, Merck, Novartis, Genzyme, Alfred Health; research support from Novartis, Biogen, Roche, Merck, NHMRC, Pennycook Foundation, MSRA. HB has received compensation for same activities from Oxford Health Policy Forum, Merck, Biogen, Novartis.

AvW has received travel support and speaker honoraria from Biogen, Sanofi Genzyme, Novartis, Teva and Merck. She receives research support from Novartis Australia and the NHMRC, Australia.

Supplemental material

Supplemental material for this article is available online.

ORCID iD

Daniel Merlo https://orcid.org/0000-0001-8766-3453

References

1. Kalmar JH, Gaudino EA, Moore NB, *et al.* The relationship between cognitive deficits and everyday functional activities in multiple sclerosis. *Neuropsychology* 2008; 22: 442–449.
2. Moccia M, Lanzillo R, Palladino R, *et al.* Cognitive impairment at diagnosis predicts 10-year multiple sclerosis progression. *Mult Scler* 2016; 22: 659–667.
3. Benedict RHB, Cookfair D, Gavett R, *et al.* Validity of the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *J Int Neuropsychol Soc* 2006; 12: 549–558.
4. Romero K, Shammi P and Feinstein A. Neurologists' accuracy in predicting cognitive impairment in multiple sclerosis. *Mult Scler Relat Disord* 2015; 4: 291–295.
5. Rocca MA, Amato MP, De Stefano N, *et al.* Clinical and imaging assessment of cognitive dysfunction in multiple sclerosis. *Lancet Neurol* 2015; 14: 302–317.
6. Langdon DW, Amato MP, Boringa J, *et al.* Recommendations for a brief international cognitive assessment for multiple sclerosis (BICAMS). *Mult Scler* 2012; 18: 891–898.
7. Benedict RH, DeLuca J, Phillips G, *et al.* Validity of the symbol digit modalities test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23: 721–733.
8. Roar M, Illes Z and Sejbaek T. Practice effect in symbol digit modalities test in multiple sclerosis patients treated with natalizumab. *Mult Scler Relat Disord* 2016; 10: 116–122.
9. Beatty WW and Goodkin DE. Screening for cognitive impairment in multiple sclerosis. An evaluation of the mini-mental state examination. *Arch Neurol* 1990; 47: 297–301.
10. Merkel B, Butzkueven H, Traboulsee AL, *et al.* Timing of high-efficacy therapy in relapsing-remitting multiple sclerosis: a systematic review. *Autoimmun Rev* 2017; 16: 658–665.
11. Kappos L, Freedman MS, Polman CH, *et al.* Long-term effect of early treatment with interferon beta-1b after a first clinical event suggestive of multiple sclerosis: 5-year active treatment extension of the phase 3 BENEFIT trial. *Lancet Neurol* 2009; 8: 987–997.
12. Benedict RH, Cohan S, Lynch SG, *et al.* Improved cognitive outcomes in patients with relapsing-remitting multiple sclerosis treated with daclizumab beta: results from the DECIDE study. *Mult Scler*. Epub ahead of print 9 May 2017. DOI: 10.1352458517707345.
13. Comi G. Effects of disease modifying treatments on cognitive dysfunction in multiple sclerosis. *Neurol Sci* 2010; 31(Suppl. 2): S261–S264.
14. Sumowski JF, Benedict R, Enzinger C, *et al.* Cognition in multiple sclerosis: state of the field and priorities for the future. *Neurology*. Epub ahead of print 17 January 2018. DOI: 10.1212/WNL.0000000000004977.
15. Silverberg NB, Ryan LM, Carrillo MC, *et al.* Assessment of cognition in early dementia. *Alzheimers Dement* 2011; 7: e60–e76.
16. Maruff P, Thomas E, Cysique L, *et al.* Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex. *Arch Clin Neuropsychol* 2009; 24: 165–178.
17. Chiaravalloti ND and DeLuca J. Cognitive impairment in multiple sclerosis. *Lancet Neurol* 2008; 7: 1139–1151.
18. Fredrickson J, Maruff P, Woodward M, *et al.* Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology* 2010; 34: 65–75.
19. Lim YY, Jaeger J, Harrington K, *et al.* Three-month stability of the CogState brief battery in healthy older adults, mild cognitive impairment, and Alzheimer's disease: results from the Australian imaging, biomarkers, and lifestyle-rate of change substudy (AIBL-ROCS). *Arch Clin Neuropsychol* 2013; 28: 320–330.
20. De Meijer L, Merlo D, Skibina O, *et al.* Monitoring cognitive change in multiple sclerosis using a computerized cognitive battery. *Mult Scler J Exp Transl Clin* 2018; 4: 205521731881551.
21. Sjonnesen K, Berzins S, Fiest KM, *et al.* Evaluation of the 9-item patient health questionnaire (PHQ-9) as an assessment instrument for symptoms of depression in patients with multiple sclerosis. *Postgrad Med* 2012; 124: 69–77.

22. Stöber J and Bittencourt J. Weekly assessment of worry: an adaptation of the Penn state worry questionnaire for monitoring changes during treatment. *Behav Res Ther* 1998; 36: 645–656.

23. Simeoni M, Auquier P, Fernandez O, *et al.* Validation of the multiple sclerosis international quality of life questionnaire. *Mult Scler* 2008; 14: 219–230.

24. Benedict RHB, Duquin JA, Jurgensen S, *et al.* Repeated assessment of neuropsychological deficits in multiple sclerosis using the symbol digit modalities test and the MS neuropsychological screening questionnaire. *Mult Scler* 2008; 14: 940–946.

25. Tombaugh T. A comprehensive review of the paced auditory serial addition test (PASAT). *Arch Clin Neuropsychol* 2006; 21: 53–76.

26. Wojcik CM, Rao SM, Schembri AJ, *et al.* Necessity of technicians for computerized neuropsychological assessment devices in multiple sclerosis. *Mult Scler*. Epub ahead of print 22 November 2018. DOI: 1352458518813287.

27. Niino M, Mifune N, Kohriyama T, *et al.* Apathy/depression, but not subjective fatigue, is related with cognitive dysfunction in patients with multiple sclerosis. *BMC Neurol* 2014; 14: 3.

28. Rao SM, Losinski G, Mourany L, *et al.* Processing speed test: validation of a self-administered, iPad(®)-based tool for screening cognitive dysfunction in a clinic setting. *Mult Scler*. Epub ahead of print 12 January 2017. DOI: 10.1177/1352458516688955.

29. Houtchens MK, Benedict RHB, Killiany R, *et al.* Thalamic atrophy and cognition in multiple sclerosis. *Neurology* 2007; 69: 1213–1223.

30. Benedict RHB, Bruce JM, Dwyer MG, *et al.* Neocortical atrophy, third ventricular width, and cognitive dysfunction in multiple sclerosis. *Arch Neurol* 2006; 63: 1301.

31. Rao SM. Role of computerized screening in healthcare teams: why computerized testing is not the death of neuropsychology. *Arch Clin Neuropsychol* 2018; 33: 375–378.

Appendix 1. One-sample *t* test of first derivative of bootstrap fitted curves ($n = 10{,}000$).

Task	Test number	First derivative of fitted curve	*n*	SD	*p* value
SRT	1	−57.09	10000	1109	2.69E−07*
	2	−7.78	10000	568	0.17
	3	3.48	10000	487	0.47
	4	−6.50	10000	626	0.30
	5	7.61	10000	638	0.23
	6	−10.02	10000	715	0.16
	7	0.88	10000	751	0.91
	8	8.92	10000	786	0.26
	9	−2.61	10000	827	0.75
	10	−7.71	10000	1611	0.63
Task	**Test number**	**First derivative of fitted curve**	***n***	**SD**	***p* value**
ChRT	1	−74.23	10000	1880	7.92E−05*
	2	−20.46	10000	754	0.007*
	3	10.86	10000	905	0.23
	4	−7.35	10000	893	0.41
	5	−0.47	10000	1009	0.96
	6	−7.15	10000	900	0.43
	7	−2.84	10000	1150	0.80
	8	14.44	10000	1637	0.38
	9	1.59	10000	1217	0.90
	10	−15.00	10000	3159	0.63
Task	**Test number**	**First derivative of fitted curve**	***n***	**SD**	***p* value**
OBK	1	−91.43	10000	3158	0.004*
	2	−49.00	10000	1198	4.34E−05*
	3	−7.32	10000	1393	0.60
	4	−7.22	10000	1299	0.58
	5	−12.44	10000	1398	0.37
	6	−18.97	10000	1536	0.22
	7	−4.36	10000	1491	0.77
	8	9.29	10000	2199	0.67
	9	0.30	10000	1812	0.99
	10	−7.33	10000	3911	0.85

*Denotes statistical significance.
ChRT, Choice Reaction Time; OBK, One-Back memory task; SRT, Simple Reaction Time; SD, standard deviation.

Appendix 2. Concordance correlation coefficient (CCC) between subsequent tests for SRT, ChRT and OBK tasks.

Between test	Task	CCC	LLCI	ULCI	SE
1 and 2	SRT	0.49	0.39	0.57	0.047
2 and 3	SRT	0.61	0.52	0.69	0.044
3 and 4	SRT	0.73	0.65	0.78	0.04
4 and 5	SRT	0.77	0.69	0.83	0.04
5 and 6	SRT	0.73	0.62	0.81	0.05
6 and 7	SRT	0.78	0.66	0.85	0.05
7 and 8	SRT	0.75	0.61	0.84	0.06
8 and 9	SRT	0.83	0.72	0.91	0.05
9 and 10	SRT	0.72	0.51	0.85	0.08
Between test	**Task**	**CCC**	**LLCI**	**ULCI**	**SE**
1 and 2	ChRT	0.56	0.47	0.64	0.04
2 and 3	ChRT	0.68	0.6	0.75	0.04
3 and 4	ChRT	0.75	0.63	0.78	0.04
4 and 5	ChRT	0.71	0.6	0.79	0.05
5 and 6	ChRT	0.78	0.68	0.85	0.04
6 and 7	ChRT	0.82	0.73	0.88	0.04
7 and 8	ChRT	0.78	0.64	0.86	0.05
8 and 9	ChRT	0.81	0.67	0.89	0.055
9 and 10	ChRT	0.73	0.51	0.86	0.09
Between test	**Task**	**CCC**	**LLCI**	**ULCI**	**SE**
1 and 2	OBK	0.68	0.62	0.74	0.03
2 and 3	OBK	0.77	0.7	0.81	0.03
3 and 4	OBK	0.84	0.78	0.88	0.02
4 and 5	OBK	0.83	0.77	0.88	0.03
5 and 6	OBK	0.86	0.8	0.9	0.025
6 and 7	OBK	0.82	0.74	0.88	0.04
7 and 8	OBK	0.85	0.76	0.9	0.04
8 and 9	OBK	0.86	0.76	0.92	0.04
9 and 10	OBK	0.9	0.8	0.95	0.034

ChRT, Choice Reaction Time; LLCI, lower-level confidence interval; OBK, One-Back memory task; ULCI, upper-level confidence interval; SE, standard error; SRT, Simple Reaction Time.