Cognitive functions over the course of 1 year in multiple sclerosis patients treated with disease modifying therapies

Kathrin S. Utz, De-Hyung Lee, Alexandra Lämmer, Anne Waschbisch, Ralf A. Linker and Thomas Schenk

Abstract

Objectives: Disease-modifying therapies (DMTs) are applied to delay or prevent disease progression in multiple sclerosis (MS). While this has mostly been proven for physical symptoms, available studies regarding long-term effects of DMTs on cognitive functions are rare and sometimes inconsistent due to methodological shortcomings. Particularly in the case of fingolimod, comprehensive data on cognitive functions are not yet available. Therefore, we set out to reliably assess cognitive functions in patients with relapsing–remitting MS (RRMS) treated with DMTs over 1 year.

Methods: Cognitive functions were assessed with eight tests at three timepoints: baseline, 6-month follow up and 12-month follow up. First, we investigated whether the stability of cognitive functions (i.e. not falling below the 5% cut-off in more than one test) over 1 year in RRMS patients ($n = 41$) corresponds to the stability in healthy individuals ($n = 40$) of a previous study. Second, we compared the percentage of declined and improved patients in the different tests. Third, we compared patients treated with fingolimod ($n = 22$) with patients treated with natalizumab ($n = 11$) with regard to cognitive stability. Fourth, based on the patient data, the Reliable Change Index was applied to compute cut-offs for reliable cognitive change.

Results: Approximately 75% of RRMS patients treated with DMTs remained stable over the course of 1 year. The Paced Auditory Serial Addition Test (PASAT) and the Spatial Recall Test (SPART), produced improvements in 12.5% and 30.6%, respectively, probably due to practice effects. Patients treated with fingolimod did not differ from patients treated with natalizumab with regard to cognitive stability.

Conclusions: Cognitive functions remain relatively stable under DMT treatment over 1 year, irrespective of the type of medication. Furthermore, the tests PASAT and SPART should be interpreted cautiously in studies examining performance changes over time. The provided RCI norms may help clinicians to determine whether a difference in two measurements observed in a RRMS patient is reliable.

Keywords: fingolimod, natalizumab, neuropsychological test, Paced Auditory Serial Addition Test, reliable change, repeated testing

Correspondence to:
Thomas Schenk, PhD
Department of Psychology,
Ludwig-Maximilians-
Universität München,
Leopoldstr. 13, 80802
München, Germany
**thomas.schenk@psy.
lmu.de**

Kathrin S. Utz, PhD
De-Hyung Lee, MD
Alexandra Lämmer, MD
Anne Waschbisch, MD
Ralf A. Linker, MD
Department of Neurology,
Friedrich-Alexander
University Erlangen-
Nuremberg, Erlangen,
Germany

Introduction

Disease progression is one of the pivotal characteristics of most multiple sclerosis (MS) subtypes. Apart from increasing physical disability, cognitive impairment may progress over time and is associated with limitations in patients' working and social life [Amato *et al.* 2001]. Neuro-axonal damage within the central nervous system is a key factor causing disability progression, also including cognitive decline [Siffrin *et al.* 2010]. In order to delay or prevent disease progression, disease-modifying therapies (DMTs) are employed, targeting some of the detrimental physiological mechanisms [Gold *et al.* 2010]. While DMTs

used to be administered parenterally, in 2010, the first new oral DMT, fingolimod, became available, that is also assumed to exert direct effects in the central nervous system [Miron *et al.* 2008; Noda *et al.* 2013]. Yet, a new drug will be only attractive if its efficacy is comparable with other DMTs. While this has been shown for fingolimod with regards a reduction in relapses and the risk of disability progression as measured by the Extended Disability Status Scale (EDSS) [Kappos *et al.* 2010], another team found no effect on EDSS progression [Calabresi *et al.* 2014], and the effects on cognitive function have not yet been comprehensively elucidated [Lovera and Kovner, 2012].

The poor evidence is in part due to the fact that studies on the effects of DMTs on cognitive function in MS patients present contradictory findings, or are difficult to interpret due to methodological shortcomings [Lovera and Kovner, 2012; Amato *et al.* 2013]. Another reason may be the lack of cognitive tests that are specifically adapted to MS patients [Utz *et al.* 2013]. Moreover, to measure cognitive change, repeated testing is required. This is challenging, because several factors may confound the assessment of true cognitive change, such as random variations, variation due to daily fluctuations of mood, motivation or alertness of patients, measurement errors, test characteristics, the test–retest reliability of a measure or the statistical phenomenon of regression towards the mean [Heaton *et al.* 2001; Benedict and Walton, 2012; Till *et al.* 2013]. Furthermore, practice effects may obscure the assessment of clinically reliable change. Notably, it became apparent that the Paced Auditory Serial Addition Test (PASAT), which is still the most widely used cognitive test in MS, is particularly prone to practice effects [Tombaugh, 2006; Utz *et al.* 2013]. Practice effects can be reduced by the use of parallel-test versions. However, for many cognitive tests, parallel versions do not exist, and practice effects may even be found when parallel versions are employed [Utz *et al.* 2013].

One solution to overcome these problems may be the computation of norms for reliable change based on healthy individuals. Another option is to determine the individual change relative to the variability observed in the group. The reliable change index (RCI) represents such an approach [Jacobson and Truax, 1991]. The RCI uses the variability in a given group to compute a cut-off

value above which changes are considered to be clinically meaningful. Chelune and colleagues adapted the RCI method to take account of potential practice effects [Chelune *et al.* 1993]. To do this, they computed the mean practice effect for their employed tests and subtracted those practice scores from the observed test scores. Heaton and colleagues concluded that the accuracy of the RCI with correction for practice effects is comparable with more complex regression models [Heaton *et al.* 2001].

So far, the RCI has only been applied in a few studies to assess cognitive changes in MS patients [Walker *et al.* 2011; Barker-Collo and Purdy, 2013; Donnchadha *et al.* 2013; Till *et al.* 2013]. In contrast to the available studies, we recruited a larger sample size ($n = 41$) compared with some studies (Walker and colleagues, $n = 12$, and Donnchadha and colleagues, $n = 26$) [Walker *et al.* 2011; Donnchadha *et al.* 2013], recruited adults, but not children or adolescents [Till *et al.* 2013], and followed the course over 12 months compared with a 6-month follow up in one of the studies [Barker-Collo and Purdy, 2013]. Furthermore, norms for reliable change based on repeated testing in healthy individuals are not routinely provided. Hence, the purpose of the current study was to assess the course of cognitive functions in patients with relapsing–remitting MS (RRMS) treated with DMTs for over 1 year.

The first aim was to examine the course of cognitive functions over 1 year in RRMS patients and compare it with the values obtained in healthy individuals examined in a previous study [Utz *et al.* 2013]. Second, we used our data to explore the course of performance in the different tests. Third, we compared RRMS patients treated with fingolimod with RRMS patients treated with natalizumab regarding the course of cognitive functions. The fourth aim was to use the RCI method to compute cut-offs for reliable cognitive change and thereby provide clinicians with norms that might allow them to assess how well individual patients respond to DMT treatment.

Methods

Patients
We recruited patients with RRMS, diagnosed according to the 2010 revised McDonald criteria, who have been in contact with the outpatient

service of the Department of Neurology of the Friedrich-Alexander University, Erlangen-Nuremberg. Patients younger than 18 years, patients who had experienced a relapse within 30 days prior to the investigation and patients whose records included a diagnosis of depression were excluded. Further exclusion criteria were dementia, drug or alcohol abuse, high-dose steroid therapy within the last 30 days, or other immune diseases requiring immune suppression. Written informed consent was obtained prior to investigation. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki II.

Cognitive tests

A psychologist administered a selection of common neuropsychological tests at three measurement timepoints (baseline, 6-month follow up and 12-month follow up). The selection of tests included subtests from the Brief Repeatable Battery of Neuropsychological Tests (BRB-N) [Rao, 1990], the 10/36 Spatial Recall Test (SPART, modified) [Utz *et al.* 2013] to assess visuospatial learning and recall, and the Paced Auditory Serial Addition Test 3' (PASAT-3') for the assessment of information-processing speed and selective attention. Furthermore, subtests from the German edition [Härting *et al.* 2000] of the Wechsler Memory Scale Revised Edition (WMS-R) [Wechsler, 1987] were used, as well as the digit span forward or backward and the spatial span forward or backward, to assess verbal or visuospatial short-term memory. Verbal long-term memory was measured via the logical memory I subtest (modified version) [Utz *et al.* 2013]. Subtests from the computerized German test battery, 'Testbatterie zur Aufmerksamkeitsprüfung' [Zimmermann and Fimm, 2009] were conducted to assess selective attention (subtest 'Go/Nogo') and divided attention (subtest 'geteilte Aufmerksamkeit').

Finally, a computer-based visual search task was applied [Utz *et al.* 2013]. Participants were required to search a visual target hidden amongst similar visual distractors. They had to indicate their choice by pointing to the target on the touch screen. Reaction (search) time (until button release), indicating visual attention, and movement time (time between button release and target touch), reflecting motor speed, were assessed. We have previously shown that the visual search task provided a more sensitive measure of MS-induced cognitive decline since it provided the best tool to discriminate between healthy individuals and MS patients. A more detailed description of the employed cognitive tests can be found in Utz and colleagues' publication [Utz *et al.* 2013].

Data analysis

All analyses were performed with IBM SPSS Statistics 19. Descriptive statistics of the patients' performance in the different tests (raw scores) at the three timepoints were determined. Data are provided for the whole patient group (*n* = 41), as well as separately for patients treated with fingolimod (*n* = 22), patients receiving natalizumab (*n* = 11) and patients treated with interferon (*n* = 7). The single patient receiving glatiramer acetate was only included in the data analysis of the whole sample, and the comparison between patients treated with fingolimod and patients treated with other therapies.

Methods for investigating the stability of cognitive functions. In order to investigate whether the stability of cognitive functions over 1 year in a sample of RRMS patients medicated with a DMT corresponds to the stability in healthy individuals, we used a previous study's data of healthy individuals [Utz *et al.* 2013]. In this study, 40 healthy individuals [12 males, 28 females; mean age, 36.3 years; standard deviation (SD): 11.39] were tested in the same tests, twice (baseline and 3–6-month follow up). The sample differed, but not significantly, with regard to age and sex distribution. Based on those data, percentile cut-off points for the difference between test scores of the 6-month follow up and baseline, and the difference between the 12-month follow up and baseline were computed. Patients falling outside the lower limit of the 5% interval were classified as cognitively declined in the respective test. There were eight tests, but some of them provided two parameters (e.g. reaction time and number of errors for Go/Nogo). In those cases, decline was diagnosed if a patient fell below the 5% cut-off in one of the parameters. This criterion was employed in the case of digit span, spatial span, SPART, and Go/Nogo.

A patient was classified as an overall 'decliner' if he or she showed significant cognitive decline in two or more of the eight tests. With a 5% error rate in each test, there is an estimated probability of 0.057 of finding decline due to chance in two or more of the eight tests. The remaining patients were classified as stable, including those whose cognitive functions improved.

http://tan.sagepub.com

271

Comparison of different tests regarding improvement and decline. For each of the cognitive tests, the percentage of patients whose performance improved and the percentage of those whose performance declined, that is, those patients whose change values either exceeded the 95% cut-offs or fell below the 5% cut-offs of healthy individuals, was determined for the 12-month follow up period. Please note that for this analysis the categories were different from those for the stability analysis (see previous section). In our analysis of cognitive stability, we combined patients whose performance remained unchanged, or whose performance improved within one category ('stable'). For the current analysis, patients whose performance remained unchanged ('stable') and those whose performance improved ('improved') were assigned to separate categories.

Analysis of the relationship between stability and type of medication. The percentage of decliners in the group of patients treated with fingolimod and patients treated with natalizumab was compared with Fisher's exact test. The same was done for the comparison between patients treated with fingolimod and patients treated with other therapies (natalizumab, interferon, glatiramer acetate). As the sample size of patients treated with interferon was small, we did not compare all three groups with one another.

Computation of norms on the basis of the Reliable Change Index. To establish norms of reliable change based on the RRMS patient sample, the Reliable Change Index (RCI) [Jacobson and Truax, 1991] was calculated for the difference between baseline and 6-month follow-up performance and the difference between baseline and 12-month follow-up performance. The RCI was computed as follows:

$$\text{Sdiff} = \sqrt{2\left(Se\right)^2},$$

where *Se* is the standard error of the measurement, computed on the basis of the SD of the patients' baseline performance in a respective test (s) and the test–retest reliability (rxx) of that test (computed here on the basis of the correlation between the patients' baseline performance and their follow-up performance):

$$Se = \text{s}\sqrt{1 - rxx}$$

We computed 90% RCI confidence intervals and corrected for practice effects [Shilling *et al.* 2005].

The practice effect for each test was the mean difference between the follow-up and baseline scores of the abovementioned sample of healthy individuals of a previous study [Utz *et al.* 2013]. However, we only introduced a correction in those tests that had shown significant practice effects in the cited study. Such practice effects were found for the PASAT-3' and digit span forward. The practice effect was 2.36 points for the PASAT-3' and 0.68 points for digit span forward. The 90% RCI intervals were computed as follows:

$$\text{RCI interval} = \left(\text{Sdiff}\right)\left(\pm 1.645\right) + \text{practice effect}$$

This formula allows the reader to compute an individual difference of the post- and pretest score of a given MS patient in a respective test and examine whether the obtained value corresponds to a meaningful change. If the computed value for the patient falls outside the limits of the confidence interval, reliable improvement or decline can be assumed (see Table 3).

Results

Patient characteristics

Initially, we recruited 73 patients, but only data of patients who did not switch medication during the study and who participated in all three measurements were included in the analysis. The final sample included 41 patients, meaning a loss of 44%. Only 22 patients (initially 42) of the sample received first-dose fingolimod and 11 patients (initially 19) received first-dose natalizumab on the day of baseline assessment. One patient was treated with glatiramer acetate (since 12 months) and seven (initially 11) with interferon beta 1a or 1b. Most interferon patients had already started treatment prior to the start of this study; one patient received his first dose at baseline. The mean duration between treatment initiation and baseline assessment of this group was 4.5 (range: 0–8) months. In the fingolimod group, 20 patients were lost to follow up; 10 of those discontinued medication during the study. In the natalizumab group eight patients were lost to follow up; one of those discontinued medication. In the interferon group four patients were lost to follow up; none of those discontinued medication.

Clinical and demographical characteristics of the entire group and subgroups are reported in Table 1. The three groups did not differ

Table 1. Patient characteristics.

	Fingolimod (n = 22)	Natalizumab (n = 11)	Interferon (n = 7)	Overall (n = 41*)
Sex: male/female	8/14	4/7	3/4	15/26
Age in years: median (range)	35 (21–60)	35 (23–51)	29 (20–48)	35 (20–60)
Time since diagnosis in months: median (range)	68 (10–300)#	28 (1–276)	24 (1–48)	42 (1–300)
EDSS				
Baseline: median (range)	2.5 (1–6) n = 22	2.5 (1–4) n = 11	1.0 (1–1.5)# n = 7	2.5 (1–6) n = 41
6-month follow up: median (range)	2.5 (1–6) n = 20	2.5 (1.5–3.5) n = 11	1.5 (0–1.5)# n = 5	2.5 (0–6) n = 37
12-month follow up: median (range)	2.5 (2–6) n = 19	2.5 (1–3.5) n = 10	1.5 (1–2.5)## n = 5	2.5 (1–6) n = 35
Visual function score within the EDSS				
Baseline: median (range)	0.0 (0–4) n = 22	0.0 (0–2) n = 10	0.0 (0–1) n = 6	0.0 (0–4) n = 39
12-month follow-up: median (range)	0.0 (0–4) n = 21	0.0 (0–2) n = 10	1.0 (0–1) n = 6	0.0 (0–4) n = 38
Functional deficit of the dominant arm (paresis, ataxia)				
Baseline, number of patients	8, n = 22	1, n = 11	0, n = 11	9, n = 41
12-month follow up, number of patients	10, n = 22	4, n = 10	0, n = 6	14, n = 39
Medication within 6 months prior to study	n = 22	n = 10	n = 6	n = 39
None	4	4	1	9
Interferon beta 1a	10	3	1	14
Interferon beta 1b	2	1	4	7
Natalizumab	3	0	0	3
Glatiramer acetate	2	2	0	5
Fingolimod	0	0	0	0
Azathioprine	1	0	0	1

* Including the patient treated with glatiramer acetate; #group differed significantly ($p < 0.05$) from the other two groups; ##group differed significantly from the fingolimod, but not the natalizumab group; EDSS, Extended Disability Status Scale. In some cases, *n* differed from the group *n* given in the first line due to missing data. Therefore, those changed *n* are indicated in the respective table cells.

significantly according to age or sex distribution (all $p > 0.05$), but differed significantly with regard to time since first diagnosis ($p = 0.019$) and EDSS score at first measurement ($p = 0.001$), at 6-month follow up ($p = 0.005$) and 12-month follow up ($p = 0.01$). Time since first diagnosis was significantly longer in patients treated with fingolimod compared with patients receiving natalizumab ($p = 0.04$) and patients treated with interferon ($p = 0.012$). Baseline EDSS score was significantly lower in patients treated with interferon than in patients treated with fingolimod ($p < .001$) and patients receiving natalizumab ($p = 0.002$). This last finding is probably not surprising given the fact that fingolimod and natalizumab were introduced as second-line DMTs. The EDSS scores of patients treated with interferon were also significantly lower than in patients treated with fingolimod at 6-month follow up ($p = 0.001$) and 12-month follow up ($p = 0.002$). Compared with the natalizumab group, patients treated with interferon had also significantly lower

EDSS scores at 6-month follow up ($p = 0.003$), whereas at 12-month follow up there was only a trend towards a lower score ($p = 0.075$) for patients treated with interferon. It should be noted that mean values for the EDSS scores remained similar across the different assessment times. However, we had to contend with more missing values for the EDSS scores at 12-month follow up and it thus seems likely that the low statistical power for this assessment time is responsible for the fact that the EDSS difference between natalizumab and interferon failed to achieve statistical significance. Groups did not differ concerning the visual function score or the frequency of a functional deficit of the right arm (all $p \geqslant 0.05$).

Patients were asked during follow-up assessment whether a relapse has occurred. Six patients reported one relapse within the 12-month study period (three of the fingolimod group, one of the natalizumab group and two of the interferon

group). One patient treated with fingolimod reported three relapses. The remaining 35 patients reported 0 relapses. None of the patients was tested during an acute relapse.

Stability of cognitive functions
Table 2 shows the mean and standard deviations of raw scores for the subgroups and the whole patient group.

Only 11 of the 41 RRMS patients (26.8 %) were classified as decliners after the 6-month follow up, whereas the majority of patients remained stable (73.2 %, including improvement). Similar results were observed after 12 months. Only 10 RRMS patients (24.4 %) showed significant cognitive decline, whereas 75.6 % remained stable or even improved in their cognitive performance. There was no significant correlation between EDSS change and cognitive change over 1 year ($r = 0.12$, $p = 0.52$).

Percentage of improved and declined patients in the different tests
For each of the cognitive tests, the percentage of improved and declined patients was determined for the 12-month follow up in order to explore the course of performance in the different cognitive measures (see Figure 1).

It is noteworthy that the percentage of improved patients is surprisingly large for the SPART and the PASAT (five times more improvers than decliners for the SPART and twice as many in the case of the PASAT). We will offer some explanation for this in the discussion section.

Relationship between stability and type of medication
In order to investigate whether the stability of cognitive function depends on the type of DMT, a Fisher's exact test was computed comparing the percentage of decliners in the groups of patients treated with fingolimod and patients treated with natalizumab.

There was no significant association between type of medication and whether or not patients showed declined or stable cognitive functions after 6 or 12 months (where $p > 0.99$ in both instances, using the two-tailed Fisher's exact test). Figure 2 shows the percentage of patients

showing either declined or stable (including improved) cognitive functions for patients treated with fingolimod ($n = 22$) and patients treated with natalizumab ($n = 11$).

As there were large differences with regard to sample size between the fingolimod and natalizumab group, we also compared the fingolimod group ($n = 22$) with the combined group of patients treated with other DMTs ($n = 19$).

There was also no significant association between type of medication and whether or not patients showed declined or stable cognitive functions after 6 or 12 months ($p > 0.05$ in both cases, using two-tailed Fisher's exact test).

Reliable change intervals as norms
To establish norms of reliable change based on the RRMS patient sample, the Reliable Change Index (RCI) was calculated for the difference between baseline and 6-month follow up performance and the difference between baseline and 12-month follow up performance. Afterwards, 90% RCI confidence intervals were computed and corrected for practice effects. Table 3 shows the reliable change intervals for each test, separately for the 6- and the 12-month follow up.

Discussion
We assessed cognitive functions over 1 year in patients with RRMS treated with DMTs. Overall, cognitive function remained relatively stable over 1 year (75.6 %). This stability did not depend on the type of medication.

There are recent studies suggesting improving or stabilizing effects of interferon beta [Lacy *et al.* 2013; Patti *et al.* 2013; Mokhber *et al.* 2014] or natalizumab [Holmen *et al.* 2011; Lang *et al.* 2012; Wilken *et al.* 2013; Iaffaldano *et al.* 2014] on cognitive functions.

Regarding natalizumab, Holmen and colleagues observed an improvement in the Symbol Digit Modalities Test in patients treated with natalizumab over 24 months [Holmen *et al.* 2011]. Lang and colleagues found stabilizing and improving effects of natalizumab on a selection of memory and attention tests, including PASAT [Lang *et al.* 2012]. Wilken and colleagues observed a stabilization or improvement in computer-based assessed cognitive functions over 1 year in patients treated

Table 2. Means and standard deviations of raw scores in the different cognitive tests.

Measure	Baseline				6-month follow up				12-month follow up			
	Fingolimod (n = 22)	Natalizumab (n = 11)	Interferon (n = 7)	Overall (n = 41*)	Fingolimod (n = 22)	Natalizumab (n = 11)	Interferon (n = 7)	Overall (n = 41*)	Fingolimod (n = 22)	Natalizumab (n = 11)	Interferon (n = 7)	Overall (n = 41*)
SPART	20.14 (5.37)	20.09 (6.12)	18.86 (5.55)	19.72 (5.53)	21.27 (5.15)	16.09 (8.13)	19.71 (5.91)	19.32 (6.58)	21.91 (5.02)	23.82 (5.04)	23.71 (4.82)	22.46 (5.23)
SPARTDR	6.95 (2.33)	6.73 (2.61)	6.14 (2.04)	6.68 (2.34)	6.82 (3.03)	6.55 (3.30)	7.71 (2.06)	6.78 (2.99)	7.64 (2.36)	7.64 (2.87)	8.43 (2.70)	7.63 (2.64)
PASAT-3'	43.10 (11.41)	40.60 (12.69)	51.57 (4.31)	45.91 (9.28)	46.19 (11.14)	44.55 (11.41)	49.57 (7.61)	48.14 (9.33)	47.62 (11.58)	49.90 (11.13)	53.57 (5.97)	51.63 (7.09)
Digit span forward	7.95 (1.94)	7.45 (1.51)	8.429 (2.23)	7.83 (1.90)	7.82 (1.87)	7.36 (1.57)	8.86 (1.95)	7.73 (2.03)	7.86 (1.78)	7.90 (1.38)	9.00 (1.41)	7.93 (1.88)
Digit span backward	6.68 (2.01)	5.99 (1.18)	6.43 (2.07)	6.44 (1.79)	7.05 (2.03)	6.18 (1.25)	7.86 (1.57)	6.83 (1.96)	7.18 (1.50)	6.82 (0.75)	7.29 (2.29)	6.98 (1.67)
Spatial span forward	8.32 (2.36)	7.55 (1.75)	8.86 (1.46)	8.1 (2.15)	9.09 (3.02)	7.45 (2.11)	9.00 (1.53)	8.51 (8.44)	8.50 (1.85)	8.45 (1.57)	8.86 (1.21)	8.44 (1.78)
Spatial span backward	7.36 (1.99)	8.18 (2.09)	9.86 (2.34)	7.95 (2.25)	7.59 (2.04)	7.27 (1.49)	7.57 (2.23)	7.41 (1.95)	7.59 (1.56)	8.36 (1.75)	8.71 (1.50)	7.88 (1.79)
Logical memory I	24.64 (10.49)	21.36 (5.37)	21.86 (6.31)	22.83 (8.96)	24.05 (8.43)	22.27 (3.00)	24.86 (7.22)	23.22 (7.60)	20.27 (6.42)	20.55 (5.41)	21.86 (2.27)	20.27 (5.94)
Go/Nogo (reaction time)	559.00 (85.08)	558.45 (54.63)	493.29 (51.96)	549.78 (76.25)	548.55 (78.10)	553.91 (61.79)	510.29 (44.31)	548.78 (76.86)	546.95 (71.17)	559.73 (79.56)	519.57 (45.72)	549.46 (72.93)
Go/Nogo (errors)	1.18 (2.59)	0.27 (0.47)	0.00 (0.00)	0.34 (1.74)	1.50 (3.36)	0.18 (0.40)	0.14 (0.38)	0.54 (1.86)	0.91 (2.97)	0.00 (0.00)	0.14 (0.38)	1.37 (5.75)
Divided attention (omissions)	3.50 (5.23)	1.91 (1.87)	1.71 (2.14)	2.93 (4.23)	2.45 (3.63)	1.55 (1.37)	1.00 (1.15)	2.12 (3.03)	1.55 (2.04)	1.36 (1.57)	0.71 (1.11)	1.73 (3.01)
Visual Search (RT)	900.73 (445.75)	863.77 (269.58)	702.17 (93.18)	862.60 (361.84)	876.09 (443.01)	813.92 (162.76)	791.10 (59.42)	875.41 (389.39)	579.32 (104.79)	538.11 (87.67)	543.60 (141.32)	983.61 (436.76)
Visual Search (MT)	745.33 (106.60)	645.99 (110.49)	706.47 (214.72)	726.62 (165.79)	705.77 (229.56)	565.33 (155.12)	556.43 (139.41)	651.07 (214.18)	1040.40 (487.67)	886.45 (123.70)	758.65 (41.823)	578.09 (148.43)

*Including the patient treated with glatiramer acetate; MT, movement time; PASAT-3', Paced Auditory Serial Addition Test 3'; RT, reaction time; SPART, Spatial Recall Test; SPARTDR, delayed recall of the Spatial Recall Test; Go/Nogo, selective-attention test.

Table 3. Reliable change intervals based on data of the whole patient sample.

Measure	6-month follow up: baseline	12-month follow up: baseline
SPART	$-9.29 \geqslant x_2 - x_1 \leqslant +9.29$	$-9.55 \geqslant x_2 - x_1 \leqslant +9.55$
SPARTDR	$-3.67 \geqslant x_2 - x_1 \leqslant +3.67$	$-3.56 \geqslant x_2 - x_1 \leqslant +3.56$
PASAT-3'	$-10.88 \geqslant x_2 - x_1 \leqslant +15.60^*$	$-10.49 \geqslant x_2 - x_1 \leqslant +15.21^*$
Digit span forward	$-1.75 \geqslant x_2 - x_1 \leqslant +3.11^*$	$-1.77 \geqslant x_2 - x_1 \leqslant +3.13^*$
Digit span backward	$-2.64 \geqslant x_2 - x_1 \leqslant +2.64$	$-2.47 \geqslant x_2 - x_1 \leqslant +2.47$
Spatial span forward	$-3.33 \geqslant x_2 - x_1 \leqslant +3.33$	$-2.71 \geqslant x_2 - x_1 \leqslant +2.71$
Spatial span backward	$-3.45 \geqslant x_2 - x_1 \leqslant +3.45$	$-3.15 \geqslant x_2 - x_1 \leqslant +3.15$
Logical memory I	$-9.77 \geqslant x_2 - x_1 \leqslant +9.77$	$-11.03 \geqslant x_2 - x_1 \leqslant +11.03$
Go/Nogo: reaction time	$-88.51 \geqslant x_2 - x_1 \leqslant +88.51$	$-98.60 \geqslant x_2 - x_1 \leqslant +98.60$
Go/Nogo: errors	$-3.49 \geqslant x_2 - x_1 \leqslant +3.49$	$-3.15 \geqslant x_2 - x_1 \leqslant +3.15$
Divided attention: omissions	$-6.81 \geqslant x_2 - x_1 \leqslant +6.81$	$-9.70 \geqslant x_2 - x_1 \leqslant +9.70$
Visual search: reaction time	$-409.80 \geqslant x_2 - x_1 \leqslant +409.80$	$-414.96 \geqslant x_2 - x_1 \leqslant +414.96$

SPART, Spatial Recall Test; SPARTDR, delayed recall of the Spatial Recall Test; PASAT-3', Paced Auditory Serial Addition Test 3'; Go/Nogo, selective-attention test; x_2, patient's follow-up test score; x_1, patient's baseline test score; *corrected for practice effects. If the computed value for the patient falls outside of the limits of the confidence interval, reliable improvement or decline can be assumed. Note that for some tests, a higher score indicates a worse, and a lower score, a better performance (Go/Nogo, divided attention and visual search), thus, reliable decline has occurred if $x_2 - x_1$ exceeds the upper limit of the interval and reliable improvement has occurred if $x_2 - x_1$ falls below the lower limit of the interval.
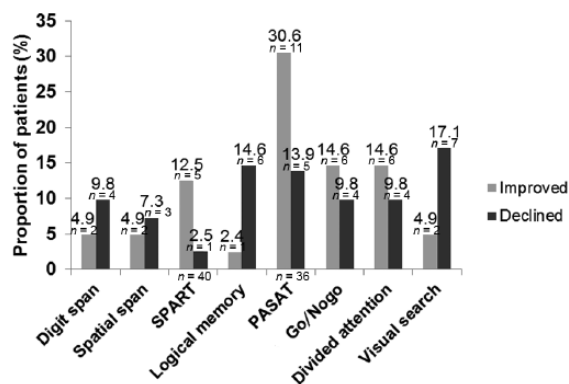

Figure 1. Proportion of improved (i.e. exceeding the 95% interval) and declined (i.e. falling outside the lower limit of the 5% interval) patients for the different tests (12-month follow up). The whole sample was considered (*n* = 41), except for SPART (one missing, because this test was included in the protocol after the first patient was examined) and PASAT (five missing due to exclusion of patients having to do the test again). The proportion of stable patients is not depicted.

with natalizumab [Wilken *et al.* 2013]. Finally, Iaffaldano and colleagues reported an improvement of cognitive functions measured with the Brief Repeatable Battery and the Stroop test over 2 years [Iffaldano *et al.* 2014]. However, since in those previous studies no data on repeated testing from healthy subjects could be provided, it is difficult to establish to what extent those stabilizing effects or improvements may have been due to practice effects.

There is no study available comprehensively investigating the effects of fingolimod on cognitive functions. There are only two studies showing stability or improvement in the Multiple Sclerosis Functional Composite (MSFC) in MS patients treated with fingolimod [Cohen *et al.* 2010; Kappos *et al.* 2010]. The MSFC is an instrument for evaluating disability in MS, including the PASAT as cognitive measure. However, no separate results for the PASAT are presented. Given that the PASAT is prone to practice effects [Tombaugh, 2006; Utz *et al.* 2013] a better performance at follow up may not necessarily indicate real cognitive improvement. The strength of our study is that we control for practice effects by defining cognitive change as exceeding the change that is expected in healthy individuals repeatedly tested in the same tests. Furthermore, we compared patients treated with fingolimod with patients treated with natalizumab and found no difference between groups. Admittedly, those results are rather exploratory, since subsample sizes are small and not perfectly matched.

When comparing the different cognitive tests regarding the percentage of improved *versus* declined patients after 12 months, the rate of improved patients strongly exceeded the rate of declined patients in the SPART and PASAT. As
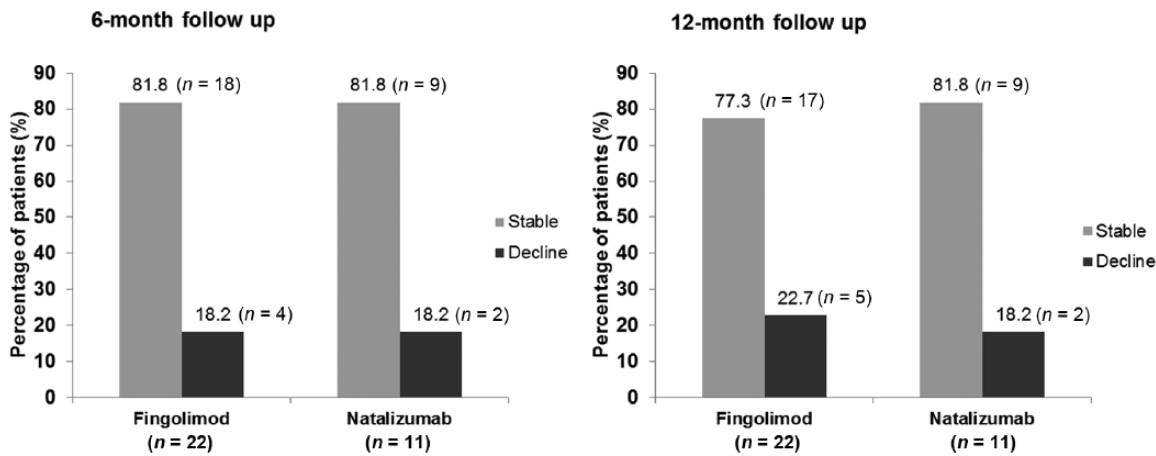
Figure 2. Percentage of patients showing either declined (i.e. falling outside the lower limit of the 5% interval in two or more tests) or stable (including improved) cognitive functions for patients treated with fingolimod (*n* = 22) and patients treated with natalizumab (*n* = 11).

in MS, more likely a decline of cognitive functions is expected, and as the percentage of decliners was higher for the other memory tests, it is plausible that the high percentage of improvement in the SPART is an artefact. Regarding the measures of attention, there was a somewhat mixed picture of more improved compared with declined patients in the selective- and divided-attention tasks, but conversely so, in the visual-search task. Nevertheless, the difference between the percentage of improved and declined patients was considerably larger for the PASAT compared with selective- and divided-attention tasks. Thus, it is very likely that the improved performance in the third test session after 12 months in those tests is in large part due to unspecific effects, such as practice effects. However, it should also be noted that while practice effects may play a role, they probably do not explain the entire extent of the observed improvements. Penner and colleagues reported stronger improvements in the PASAT in patients treated with interferon beta-1b as compared with an untreated control group [Penner *et al.* 2012]. Clearly, practice effects cannot solely explain these findings.

Nevertheless, substantial practice effects have already been reported for the PASAT before [Tombaugh, 2006; Utz *et al.* 2013] and the question remains as to why the figures for PASAT and the SPART suggest an increased susceptibility for practice effects despite our best efforts to control for practice effects on the basis of data from healthy control participants. One explanation seems to be that practice effects are more pronounced for MS

patients as compared with healthy controls. It is known that the magnitude of practice effects is influenced by the characteristics of the subjects and tests. They are, for instance, larger in younger subjects or in tests involving problem solving and a great degree of novelty [Heaton *et al.* 2001]. It has been hypothesized that improved cognitive performance during DMT treatment may be due to enhancement of neural networks involved in practice effects [Lovera and Kovner, 2012]. Furthermore, it is noteworthy that five patients refused to do the PASAT again at follow-up testing because they experienced it as very stressful. This observation is in line with reports in the literature about the arousal of frustration and aversion on the patient's side caused by PASAT [Tombaugh, 2006]. To sum up, our results contribute to the increasing evidence that the PASAT is not suitable for assessing the course of cognitive functions [Tombaugh, 2006; Utz *et al.* 2013].

The same seems to be the case for the SPART that is like the PASAT part of the BRB-N (Rao, 1990). Thus, healthy individuals may not be the ideal norm group for cognitive function in MS patients, as practice effects may be stronger in the latter group and thus mask cognitive decline. For this reason, we recommend the use of the confidence intervals based on the RRMS sample. These intervals give an indication of the amount of change that is expected and thus, conversely, can also give an indication of when a given change falls outside the expected range, and thus constitutes evidence of either meaningful improvement or meaningful decline.

The following potential limitations to our study should be noted. First, the period of 1 year is relatively short. Possibly, differences in cognitive functions between patient groups will only be detectable after a longer treatment period. Second, we did not present MRI data that would have allowed the comparison of brain lesions with change in cognitive functions. Third, the overall sample size is relatively small and only RRMS patients were included, thus limiting our ability to generalize our findings to the MS population at large. Also the fact that about 44% of the initial sample was not included due to discontinuation of medication or loss to follow up may contribute to this limitation. Especially in the fingolimod group, half of the patients who were lost to the follow-up study had discontinued with their medication. We cannot rule out the possibility that this significant dropout rate may have led to some distortion of the results. Furthermore, the patient subgroups were not perfectly matched regarding sample size and baseline characteristics. Time since first diagnosis was significantly longer in patients treated with fingolimod compared with the other groups, and baseline EDSS scores were significantly lower in patients treated with interferon than in the other groups. This last finding is probably not surprising since interferon is typically the first DMT offered, whereas the other two drugs are typically offered after treatment with another DMT proved unsatisfactory.

The fact that fingolimod patients show longer disease duration but similar age and EDSS scores compared with patients treated with natalizumab might imply a milder disease course (i.e. earlier onset, but similar EDSS). It is likely that this is the result of a self-selection process. When the study started, fingolimod was new and used as second-line DMT. Possibly, patients suffering longer from the disease and thus having more experiences with DMTs were more likely to choose an oral DMT as a second-line DMT. In contrast, patients with a more accelerated disease course may have opted for natalizumab as the more established second-line DMT. This might explain the difference in disease duration for patients on fingolimod *versus* patients on natalizumab. Furthermore, whereas patients treated with fingolimod or natalizumab received their first dose at baseline assessment, therapy of patients in the interferon group had started about 4.5 months before. Therefore, interferon had longer time to affect cognitive functions compared with the other two DMTs, which might have influenced the results. This is particularly relevant for the comparison between the groups.

Unfortunately, our study did not include an untreated patient control group. Such a group could have provided us with a baseline for the rate of change expected in the absence of treatment and thus would have allowed us to estimate more precisely the extent to which DMTs can modify the cognitive change expected in MS patients. However, an untreated control group would have been difficult to recruit, as typically, patients come to our centre to receive treatment. Denying treatment to patients who wanted to be treated would have been ethically unacceptable in the context of our study. Moreover, recruiting only those patients who themselves were not interested in receiving treatment would have resulted in a small and unrepresentative sample. Without such a control patient group we can only make statements about the cognitive stability that is observed in DMT-treated patients. Nevertheless, such information can be quite valuable since it provides a yardstick against which the response of individual patients to DMT can be judged.

In conclusion, we found that a majority (approximately 75%) of RRMS patients treated with DMTs remain stable over the course of 1 year. The specific type of DMT applied did not affect this finding. We also found that two of the most frequently used cognitive tests (PASAT and SPART) produced improvements in a surprisingly large proportion of patients. We assume that these improvements are most likely indications of practice effects and not markers of amelioration. Accordingly, these two tests should be treated with caution in studies examining performance changes over time. Finally, our study yielded confidence intervals that can be used to assess whether an observed cognitive change in a given MS patient should be seen as reliable sign of change or can be discounted as expected performance fluctuation.

Conflict of interest statement

References

Amato, M., Langdon, D., Montalban, X., Benedict, R., Deluca, J., Krupp, L. *et al.* (2013) Treatment of cognitive impairment in multiple sclerosis: position paper. *J Neurol* 260: 1452–1468.

Amato, M., Ponziani, G., Siracusa, G. and Sorbi, S. (2001) Cognitive dysfunction in early-onset multiple sclerosis: a reappraisal after 10 years. *Archives of Neurology* 58: 1602–1606.

Barker-Collo, S. and Purdy, S. (2013) Determining the presence of reliable change over time in multiple sclerosis: evidence from the PASAT, adjusting-PSAT, and Stroop test. *Int J MS Care* 15: 170–178.

Benedict, R. and Walton, M. (2012) Evaluating cognitive outcome measures for MS clinical trials: what is a clinically meaningful change? *Mult Scler* 18: 1673–1679.

Calabresi, P., Radue, E., Goodin, D., Jeffery, D., Rammohan, K., Reder, A. *et al.* (2014) Safety and efficacy of fingolimod in patients with relapsing–remitting multiple sclerosis (FREEDOMS II): a double-blind, randomised, placebo-controlled, phase III trial. *The Lancet Neurology* 13: 545–556.

Chelune, G., Naugle, R., Lüders, H., Sedlak, J. and Awad, I. (1993) Individual change after epilepsy surgery: practice effects and base-rate information. *Neuropsychology* 7: 41–52.

Cohen, J., Barkhof, F., Comi, G., Hartung, H., Khatri, B., Montalban, X. *et al.* (2010) Oral fingolimod or intramuscular interferon for relapsing multiple sclerosis. *N Engl J Med* 362: 402–415.

Donnchadha, S., Burke, T., Bramham, J., O'Brien, M., Reilly, R., Kiiski, H. *et al.* (2013) Reliable change

indices for the brief international cognitive assessment for multiple sclerosis (BICAMS). *Multiple Sclerosis Journal* 19: 971.

Gold, R., Wolinsky, J., Amato, M. and Comi, G. (2010) Evolving expectations around early management of multiple sclerosis. *Ther Adv Neurol Disord* 3: 351–367.

Härting, C., Markowitsch, H., Neufeld, H., Calabrese, P., Deisinger, K. and Kessler, J. (2000) *Wechsler gedächtnistest – revidierte fassung*. Bern: Huber.

Heaton, R., Temkin, N., Dikmen, S., Avitable, N., Taylor, M., Marcotte, T. *et al.* (2001) Detecting change: a comparison of three neuropsychological methods, using normal and clinical samples. *Arch Clin Neuropsychol* 16: 75–91.

Holmen, C., Piehl, F., Hillert, J., Fogdell-Hahn, A., Lundkvist, M., Karlberg, E. *et al.* (2011) A Swedish national post-marketing surveillance study of natalizumab treatment in multiple sclerosis. *Mult Scler* 17: 708–719.

Iaffaldano, P., Ruggieri, M., Viterbo, R., Mastrapasqua, M. and Trojano, M. (2014) The improvement of cognitive functions is associated with a decrease of plasma osteopontin levels in natalizumab treated relapsing multiple sclerosis. *Brain Behavior and Immunity* 35: 176–181.

Jacobson, N. and Truax, P. (1991) Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 59: 12–19.

Kappos, L., Radue, E., O'Connor, P., Polman, C., Hohlfeld, R., Calabresi, P. *et al.* (2010) A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis. *N Engl J Med* 362: 387–401.

Lacy, M., Hauser, M., Pliskin, N., Assuras, S., Valentine, M. and Reder, A. (2013) The effects of long-term interferon-beta-1b treatment on cognitive functioning in multiple sclerosis: a 16-year longitudinal study. *Mult Scler* 19: 1765–1772.

Lang, C., Reiss, C. and Maurer, M. (2012) Natalizumab may improve cognition and mood in multiple sclerosis. *Eur Neurol* 67: 162–166.

Lovera, J. and Kovner, B. (2012) Cognitive impairment in multiple sclerosis. *Curr Neurol Neurosci Rep* 12: 618–627.

Miron, V., Schubart, A. and Antel, J. (2008) Central nervous system-directed effects of FTY720 (Fingolimod). *Journal of the Neurological Sciences* 274: 13–17.

Mokhber, N., Azarpazhooh, A., Orouji, E., Rao, S., Khorram, B., Sahraian, M. *et al.* (2014) Cognitive dysfunction in patients with multiple sclerosis treated with different types of interferon beta: a randomized clinical trial. *J Neurol Sci* 342: 16–20.

Noda, H., Takeuchi, H., Mizuno, T. and Suzumura, A. (2013) Fingolimod phosphate promotes the neuroprotective effects of microglia. *J Neuroimmunol* 256: 13–18.

Patti, F., Morra, V., Amato, M., Trojano, M., Bastianello, S., Tola, M. *et al.* (2013) Subcutaneous interferon beta-1a may protect against cognitive impairment in patients with relapsing–remitting multiple sclerosis: 5-year follow-up of the COGIMUS study. *PLoS ONE* 8: e74111.

Penner, I., Stemper, B., Calabrese, P., Freedman, M., Polman, C., Edan, G. *et al.* (2012) Effects of interferon beta-1b on cognitive performance in patients with a first event suggestive of multiple sclerosis. *Mult Scler* 18: 1466–1471.

Rao, S. (1990) *Cognitive function study group of the national multiple sclerosis society. A manual for the brief repeatable battery of neuropsychological tests in multiple sclerosis*. Milwaukee: Medical College of Wisconsin.

Shilling, V., Jenkins, V., Morris, R., Deutsch, G. and Bloomfield, D. (2005) The effects of adjuvant chemotherapy on cognition in women with breast cancer–preliminary results of an observational longitudinal study. *Breast* 14: 142–150.

Siffrin, V., Vogt, J., Radbruch, H., Nitsch, R. and Zipp, F. (2010) Multiple sclerosis - candidate mechanisms underlying CNS atrophy. *Trends Neurosci* 33: 202–210.

Till, C., Racine, N., Araujo, D., Narayanan, S., Collins, D., Aubert-Broche, B. *et al.* (2013) Changes in cognitive performance over a 1-year period in children and adolescents with multiple sclerosis. *Neuropsychology* 27: 210–219.

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

Utz, K., Hankeln, T., Jung, L., Lammer, A., Waschbisch, A., Lee, D. *et al.* (2013) Visual search as a tool for a quick and reliable assessment of cognitive functions in patients with multiple sclerosis. *PLoS ONE* 8: e81531.

Walker, L., Mendella, P., Stewart, A., Freedman, M. and Smith, A. (2011) Meaningful change in cognition in multiple sclerosis: method matters. *Can J Neurol Sci* 38: 282–288.

Wechsler, D. (1987) *WMS-R: Wechsler Memory Scale - revised* (Manual). San Antonio: The Psychological Corporation.

Wilken, J., Kane, R., Sullivan, C., Gudesblatt, M., Lucas, S., Fallis, R. *et al.* (2013) Changes in fatigue and cognition in patients with relapsing forms of multiple sclerosis treated with natalizumab: the Ener-G Study. *Int J MS Care* 15: 120–128.

Zimmermann, P. and Fimm, B. (2009) *Testbatterie zur aufmerksamkeitsprüfung, 2.2 vol.* Herzogenrath: Psytest.