Longitudinal Changes in Cognitive Test Scores in Patients With Relapsing-Remitting Multiple Sclerosis

An Analysis of the DECIDE Dataset

Nina Castrogiovanni, BSc, Jop Mostert, MD, PhD, Pavle Repovic, MD, PhD, James D. Bowen, MD,
Bernard M.J. Uitdehaag, MD, PhD, Eva M.M. Strijbis, MD, PhD, Gary R. Cutter, PhD, and
Marcus W. Koch, MD, PhD

Correspondence
Dr. Koch
mwkoch@ucalgary.ca

Abstract

Background and Objectives

Cognitive impairment is a common and impactful symptom of relapsing-remitting multiple sclerosis (RRMS). Cognitive outcome measures are often used in cross-sectional studies, but their performance as longitudinal outcome measures in clinical trials is not widely researched. In this study, we used data from a large clinical trial to describe change on the Symbol Digit Modalities Test (SDMT) and the Paced Auditory Serial Addition Test (PASAT) over up to 144 weeks of follow-up.

Methods

We used the data set from DECIDE (clinicaltrials.gov identifier NCT01064401), a large randomized controlled RRMS trial to describe change on the SDMT and PASAT over 144 weeks of follow-up. We compared change on these cognitive outcomes with change on the timed 25-foot walk (T25FW), a well-established physical outcome measure. We investigated several definitions for clinically meaningful change: any change, 4-point change, 8-point change, and 20% change for the SDMT, any change, 4-point change, and 20% change for the PASAT, and 20% change for the T25FW.

Results

DECIDE included 1,814 trial participants. SDMT and PASAT scores steadily improved throughout follow-up: the SDMT from a mean 48.2 (SD, 16.1) points at baseline to 52.6 (SD 15.2) at 144 weeks and the PASAT from 47.0 (SD 11.3) at baseline to 50.0 (SD 10.8) at 144 weeks. This improvement in scores is most likely due to a practice effect. Throughout the trial, participants were more likely to experience improvement than worsening of their SDMT and PASAT performance, whereas the number of worsening events on the T25FW steadily increased. Changing the definition of clinically meaningful change for the SDMT and PASAT or using a 6-month confirmation changed the overall number of worsening or improvement events but did not affect the overall behavior of these measures.

Discussion

Our findings suggest that the SDMT and PASAT scores do not accurately reflect the steady cognitive decline that people with RRMS experience. Both outcomes show postbaseline increases in scores, which complicates the interpretation of these outcome measures in clinical trials. More research into the size of these changes is needed before recommending a general threshold for clinically meaningful longitudinal change.

MORE ONLINE

CME Course
NPub.org/cmelist

From the Department of Clinical Neurosciences (N.C., M.W.K.), University of Calgary, Canada; Department of Neurology (J.M.), Rijnstate Hospital, Arnhem, The Netherlands; Multiple Sclerosis Center (P.R., J.D.B.), Swedish Neuroscience Institute, Seattle, WA; Department of Neurology (B.M.J.U., E.M.M.S.), MS Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, The Netherlands; Department of Biostatistics (G.R.C.), University of Alabama at Birmingham; Department of Community Health Sciences (M.W.K.), University of Calgary, Canada.

Glossary

CC = correlation coefficient; **CELs** = contrast-enhancing lesions; **EDSS** = Expanded Disability Status Scale; **IQR** = interquartile range; **MS** = multiple sclerosis; **PASAT** = Paced Auditory Serial Addition Test; **RRMS** = relapsing-remitting multiple sclerosis; **SDMT** = Symbol Digit Modalities Test; **T25FW** = timed 25-foot walk.

Cognition is affected in all forms of multiple sclerosis (MS), and cognitive dysfunction in MS can occur from disease onset, including at the clinically isolated syndrome stage.[1] Cognitive dysfunction in adults with MS is common; its prevalence depends on the definition of impairment and cognitive tests used and varies between 43% and 65%.[2]

Clinical trials in all forms of MS often include measures of cognitive function, but most studies examining cognition in MS are cross-sectional. Cross-sectional analyses reflect differences in outcome across the entire spectrum of disease duration and participant ages, whereas longitudinal studies describe change over a range of only a few years. It is unclear whether cognitive measures as currently assessed are useful longitudinal outcomes over the two- or three-year duration of a typical clinical trial. In addition, measures taken at a single time point may fail to identify patients with progressive cognitive decline who have yet to reach a threshold defined as clinically meaningful. Two commonly used cognitive trial outcomes are the 3-second Paced Auditory Serial Addition Test (PASAT) and the Symbol Digit Modalities Test (SDMT). The PASAT is a measure of processing speed and executive function and requires patients to perform a series of mental additions under time pressure. The PASAT is often experienced as stressful,[3] and therefore, many trial participants do not tolerate it well. The PASAT also has a substantial practice effect, which limits its usefulness as a trial outcome and makes it difficult to estimate change over the trial period.[4] The SDMT was originally developed as a screening test[5] and predominantly measures processing speed. The SDMT requires trial participants to match symbols to numbers in a set period of time. The SDMT is generally seen as more patient-friendly and was recommended as the standard longitudinal outcome for clinical trials in MS.[6-8] Initially, a 4-point decrease in SDMT performance was understood to be clinically relevant as this cutoff is associated with the loss of employment[9]; more recently, an 8-point decrease has been proposed as a reliable threshold of cognitive decline.[10] Validated cutoff scores do not exist for the PASAT.

In an analysis of a large clinical trial data set, we found that people with secondary progressive MS showed a steady improvement on the SDMT throughout 2 years of follow-up.[11] This improvement in SDMT scores occurred in contrast to steady worsening on physical outcome measures and may be due to a practice effect. Other possibilities include a placebo effect and the net effects of selective dropouts. In this investigation, we use patient-level data from DECIDE, a large phase 3 randomized controlled trial in relapsing-remitting MS (RRMS), to study change in SDMT and PASAT scores over 3 years of follow-up.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

The DECIDE data set was obtained from Biogen (Biogen, Cambridge, MA), the pharmaceutical company that conducted and oversaw this trial. Ethical approval for the original trial is described in the original publication.[12] Ethical approval for this analysis was sought and granted by the University of Calgary Conjoint Health Research Ethics Board.

Data Set

DECIDE was a randomized controlled phase 3 study comparing subcutaneous daclizumab, administered every 4 weeks, with IM interferon beta-1a, administered once weekly, in people with RRMS for up to 144 weeks. Entry criteria included a confirmed diagnosis of RRMS, age 18–55 years, a score of 0–5.0 inclusive on the Expanded Disability Status Scale (EDSS),[13] and clinical relapses in the 2 years before inclusion.

Statistical Analyses

In DECIDE, the SDMT was measured every 24 weeks, and the PASAT was measured every 12 weeks. As a measure of physical disability, we chose the Timed 25-foot Walk (T25FW),[14] which we previously showed to be a reliable and robust measure of physical disability in RRMS.[15] The T25FW was performed every 12 weeks in DECIDE. For this analysis, we focused on the measurements at baseline, 24, 48, 72, 96, 120, and 144 weeks.

We determined the proportion of participants with unconfirmed and 6-month confirmed worsening and improvement on these outcome measures by comparing baseline and follow-up measures. Individuals missing a score at baseline, the time point of interest, or the corresponding 6-month confirmation assessment were excluded from the analysis.

We explored different definitions for clinically meaningful change for the SDMT and PASAT. For the SDMT we used (1) any change from baseline, (2) the established threshold[9] of a 4-point change from baseline, (3) the recently suggested[10] threshold of an 8-point change from baseline, and (4) 20% or more change from baseline. For the PASAT, we used (1) any change from baseline, (2) 4-point change from baseline, and (3) 20% or more change from baseline. For the T25FW, we defined clinically meaningful change as 20% or more change from baseline.[16]

We compared cognitive outcome scores and change in cognitive outcome scores between the treatment arms with Student t tests. Statistical significance was taken to be at the two-sided 0.05 level. We used the R statistical software package for Windows version 4.2.2 for all statistical analyses.[17]

Change in Cognitive Scores by Cognitive Performance at Baseline

Because the change in cognitive outcomes could vary depending on the degree of cognitive dysfunction at baseline,[18] we investigated the mean change from baseline in SDMT and PASAT scores in individuals with below-average vs average-and-above scores at baseline. For these analyses, we used the average scores at baseline in the DECIDE data set as cutoffs: 49 for the SDMT and 50 for the PASAT (Table 1).

Baseline Factors Associated With Change in Cognitive Scores

To investigate which baseline patient factors may be associated with the change in cognitive scores, we built linear regression models with the change in SDMT and PASAT scores between baseline and 96 weeks as the dependent (outcome) variable and age, sex, disease duration, treatment arm, and the cognitive score at baseline as independent (predictor) variables.

Estimate of Regression to the Mean

To illustrate the possible influence of regression to the mean on our results, we first determined the Pearson correlation coefficient (CC) of the first (baseline) and second (24 weeks) SDMT and PASAT measurements. We then calculated Z-scores for the first and second SDMT and PASAT measurements and selected cases with a baseline Z-score ≤ -1.282, which represent the lowest 10% of normally distributed test scores. We then calculated the mean Z-score for the 24-week measurement in these selected cases. We used the

following formula to estimate the size of the effect of regression to the mean in the selected cases:

$$(1 - CC) * (mean\ Z\text{-}score\ 24\ weeks - mean\ Z\text{-}score\ baseline)$$

We multiplied the result of this formula with the SD of the baseline measurement in the whole cohort to arrive at an estimate of regression to the mean in SDMT or PASAT units.

Data Availability

The data used in this study are available on request from Biogen. Individual participant data collected during the trial will be shared after anonymization and on approval of a research proposal and data sharing agreement. Research proposals can be submitted online (biogenclinicaldatarequest.com).

Results

DECIDE Data Set

The DECIDE data set contained patient-level data from 1,814 trial participants. Their baseline characteristics are given in Table 1. DECIDE was designed to follow the participant cohort until the last participant had finished the 96-week visit to a maximum of 144 weeks. As a consequence of this design, there were substantially fewer participants at the 120- and 144-week follow-up visits. At 96 weeks, there were 1,311 participants (72.2% of the baseline cohort) with SDMT measurements in the cohort, compared to 699 (38.5% of the baseline cohort) at 120 weeks and 383 (21.1% of the baseline cohort) at 144 weeks. For the PASAT, there were 1,316 (72.5% of the baseline cohort) measurements at 96 weeks, 703 (38.8% of the baseline cohort) at 120 weeks, and 388 (21.4% of the baseline cohort) at 144 weeks (Table 2).

Change in Cognitive Outcome Scores

Mean SDMT and PASAT scores steadily increased during follow-up (Figure 1 and Table 2). At 144 weeks of follow-up,

Table 1 Baseline Characteristics

	Interferon beta-1a	Daclizumab	Total
n	922	919	1,841
Sex (F/M)	627/295	625/294	1,252/589
Age (mean, SD)	36.2, 9.3	36.4, 9.3	36.3, 9.3
Disease duration (mean, SD)	6.9, 6.3	7.0, 6.3	6.9, 6.3
Relapses in the year before inclusion (mean, SD)	1.6, 0.8	1.5, 0.7	1.6, 0.7
Relapses in the 2 y before inclusion (mean, SD)	2.3, 1.0	2.2, 0.9	2.2, 1.0
Participants with CELs at baseline (%)	45.3%	44.0%	44.7%
EDSS score at baseline (median, IQR)	2.5 (1.5–3.5)	2.0 (1.5–3.5)	2.0 (1.5–3.5)
SDMT at baseline (median, IQR)	49 (37–58)	49 (39–58)	49 (38–58)
PASAT at baseline (median, IQR)	50 (40–56)	51 (41–56)	50 (40–56)

Abbreviations: CELs = contrast-enhancing lesions; EDSS = Expanded Disability Status Scale; IQR = interquartile range; PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test.

Table 2 Raw Outcome Scores Throughout Follow-up

	Week of follow-up						
	Baseline	24	48	72	96	120	144
SDMT (mean, SD)	48.2, 16.1	49.1, 15.3	50.5, 15.4	51.6, 15.1	51.7, 15.1	52.3, 15.2	52.6, 15.2
SDMT (median, IQR)	49, 38 to 58	50, 39 to 59	51, 41 to 61	52, 42 to 62	52, 42 to 62	52, 43 to 62	53, 43 to 63
Number of observations	1,775	1,678	1,549	1,425	1,311	699	383
SDMT change from baseline (mean, SD)	—	1.1, 11.6	2.4, 12.0	3.3, 12.7	3.8, 12.4	4.7, 12.8	4.9, 12.5
SDMT change from baseline (median, IQR)	—	2, −3 to 6	2, −3 to 8	3, −2 to 9	4, −2 to 10	4, −2 to 11	5, −1 to 11
Number of observations	—	1,630	1,504	1,386	1,276	683	370
PASAT (mean, SD)	47.0, 11.3	48.2, 10.8	49.3, 10.4	49.9, 10.4	50.0, 10.6	50.2, 10.4	50.0, 10.8
PASAT (median, IQR)	50, 40 to 56	51, 43 to 57	53, 44 to 57	54, 45 to 58	54, 45 to 58	54, 45 to 58	54, 44.47 to 58.25
Number of observations	1,827	1,693	1,516	1,439	1,316	703	388
PASAT change from baseline (mean, SD)	—	1.2, 6.3	2.0, 7.1	2.5, 7.1	2.6, 7.5	3.1, 7.6	3.3, 8.1
PASAT change from baseline (median, IQR)	—	1, −2 to 4	1, −1 to 6	2, −1 to 6	2, −1 to 6	2, −1 to 7	2, −1 to 7
Number of observations	—	1,684	1,552	1,429	1,310	699	385

Abbreviations: IQR = interquartile range; PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test.

mean SDMT scores had increased by 4.9 points (SD 12.5), and mean PASAT scores had increased by 3.3 points (SD 8.1) compared with baseline.

There were significant differences in SDMT and PASAT scores between treatment arms from baseline to 144 weeks of follow-up, with significantly higher scores in the daclizumab arm (Figure 1 and Table 3). In both treatment arms, SDMT and PASAT scores increased over time, significantly more in the daclizumab arm than in the interferon arm: at 144 weeks, mean SDMT scores increased by 6.8 (SD 12.2) in the daclizumab arm, compared with 2.9 (SD 12.5) in the interferon arm; however, at 48 weeks, the significant differences were in the opposite direction with the interferon beta-1a arm showing greater change. At 144 weeks, mean PASAT scores increased by 4.0 points (SD 8.1) in the daclizumab arm, compared with 2.6 points (SD 8.1) in the interferon arm (Table 3). Although these differences were statistically

Figure 1 Mean Cognitive Outcome Scores During Follow-up
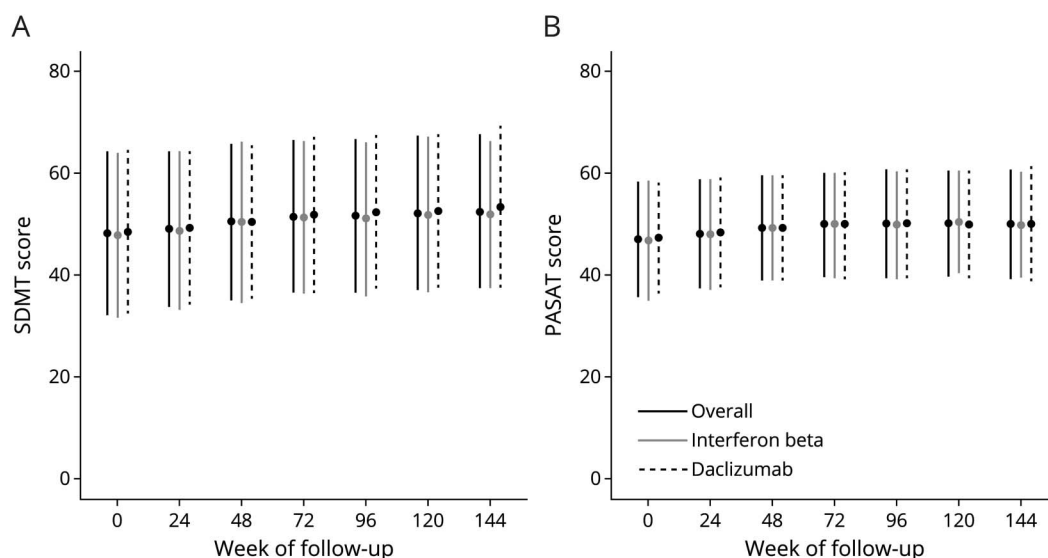
Mean SDMT (A) and PASAT (B) scores throughout follow-up. Error bars represent the SD. Mean scores on both instruments steadily increase throughout follow-up, likely due to a practice effect of the repeated application of the measures. There is little difference between the treatment arms in cognitive scores.

Table 3 Mean Outcome Scores Throughout Follow-up in the 2 Treatment Arms

		Week of follow-up						
		Baseline	24	48	72	96	120	144
Daclizumab	Mean SDMT, SD	48.6, 16.0	49.3, 15.1	50.5, 15.0	51.9, 15.3	52.4, 15.1	52.7, 15.0	53.4, 15.9
	n (% of enrolled)	889 (96.7)	856 (93.1)	799 (86.9)	729 (79.3)	667 (72.6)	352 (38.3)	192 (20.9)
Interferon	Mean SDMT (SD)	47.8, 16.2	48.8, 15.6	50.4, 15.8	51.3, 14.9	51.0, 15.1	51.9, 15.3	51.9, 14.5
	n (% of enrolled)	886 (96.1)	822 (89.2)	750 (81.3)	696 (75.5)	644 (69.8)	347 (37.6)	191 (20.7)
	p (*t* test)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Daclizumab	SDMT mean change from baseline, SD	—	1.1, 12.3	2.2, 11.7	3.7, 12.4	4.7, 11.8	5.7, 12.6	6.8, 12.2
	n (% of enrolled)	—	834 (90.8)	779 (84.8)	709 (77.1)	650 (70.7)	345 (37.5)	185 (20.1)
Interferon	SDMT mean change from baseline, SD	—	1.2, 10.9	2.5, 12.4	3.0, 13.0	2.8, 13.0	3.7, 13.0	2.9, 12.5
	n (% of enrolled)	—	796 (86.3)	725 (78.6)	677 (73.4)	626 (67.9)	345 (37.4)	185 (20.1)
	p (*t* test)	—	0.17	**0.0005**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Daclizumab	Mean PASAT, SD	47.3, 10.9	48.4, 10.7	49.3, 10.4	49.8, 10.6	50.2, 10.6	50.0, 10.7	50.0, 11.4
	n (% of enrolled)	911 (99.1)	866 (94.2)	806 (87.7)	739 (80.4)	671 (73.0)	353 (38.4)	194 (21.1)
Interferon	Mean PASAT, SD	46.8, 11.7	48.0, 10.9	49.3, 10.3	50.0, 10.3	49.8, 10.6	50.5, 10.1	49.9, 10.4
	n (% of enrolled)	916 (99.3)	827 (89.7)	756 (82.0)	700 (75.9)	645 (70.0)	350 (38.0)	194 (21.0)
	p (*t* test)	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
Daclizumab	PASAT mean change from baseline, SD	—	1.2, 5.8	2.1, 6.7	2.4, 6.9	2.8, 7.3	3.5, 7.6	4.0, 8.1
	n (% of enrolled)	—	861 (93.7)	800 (87.1)	733 (79.8)	668 (72.7)	351 (38.2)	192 (20.9)
Interferon	PASAT mean change from baseline, SD	—	1.1, 6.7	2, 7.4	2.6, 7.3	2.4, 7.8	2.7, 7.6	2.6, 8.1
	n (% of enrolled)	—	823 (89.3)	752 (81.6)	696 (75.5)	642 (69.6)	348 (37.7)	193 (20.9)
	p (*t* test)	—	**0.02**	**0.002**	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**

Abbreviations: PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test.
Bold type indicates statistically significant *p* values.

significant, the overall difference in raw scores was low, for example, at 96 weeks of follow-up, the mean SDMT score was 52.4 (SD 15.1) in the daclizumab arm, compared with 51.0 (SD 15.1) in the interferon arm. Similarly, the mean PASAT score at 96 weeks of follow-up was 50.2 (SD 10.6) in the daclizumab arm and 49.8 (SD 10.6) in the interferon arm (Figure 1 and Table 3).

### Clinically Meaningful Change in Cognitive Outcomes

Throughout the trial, participants were more likely to experience improvement than worsening of their SDMT and PASAT performance, and improvement events increased steadily over follow-up, whereas there was little change in worsening events for the SDMT and PASAT (Figure 2 and Table 4). These findings for cognitive outcomes stand in contrast to change on the T25FW, a physical outcome measure. On the T25FW, the number of worsening events steadily increased over time, and

worsening events always outnumbered improvement events. Using 6-month confirmation yielded lower estimates than for unconfirmed change but did not alter the overall behavior of the measures (Table 4). Changing the definition of clinically meaningful change for the SDMT and PASAT similarly changes the overall number of worsening or improvement events but does not affect the overall behavior of these measures (Table 4).

Change in Cognitive Scores by Cognitive Performance at Baseline

SDMT and PASAT scores increased in participants with below-average cognitive performance and in those with average-and-above cognitive performance at baseline (Table 5). Scores increased more in participants with below-average performance at baseline for both measures (mean change from baseline at 96 weeks: 7.3 SDMT points, 5.2 PASAT points) compared with those with average-and-above performance at baseline (mean

Figure 2 Worsening and Improvement Events During Follow-up
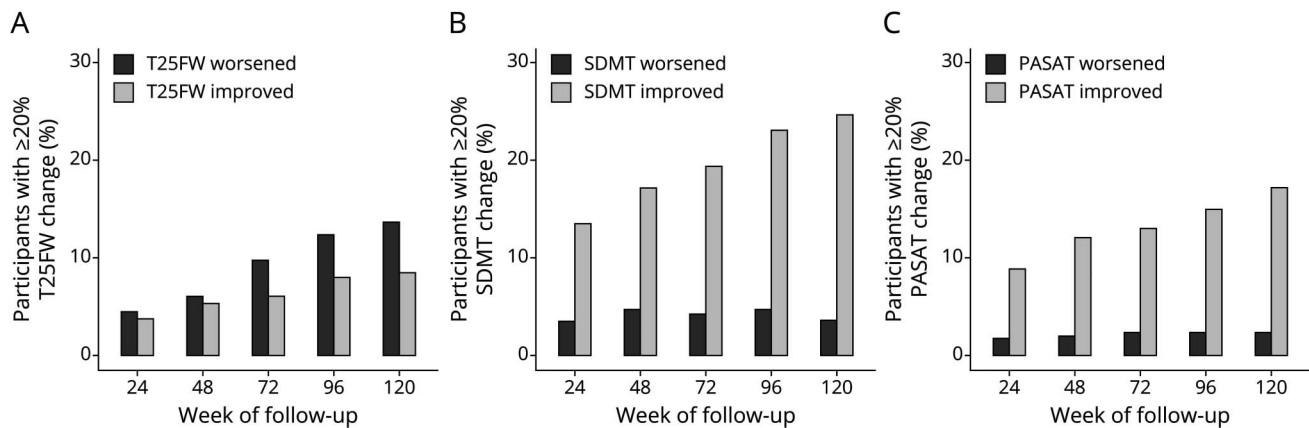
Participants with 6-month confirmed ≥20% worsening or improvement on the T25FW (A), SDMT (B), and PASAT (C). As expected, worsening events on the T25FW steadily increase throughout the trial, and worsening events always outnumber improvement events. In contrast, on the SDMT and PASAT, worsening events remain low and stable throughout follow-up, whereas improvement events steadily increase. Improvement events always outnumber worsening events on the SDMT and PASAT.

change from baseline at 96 weeks: 0.4 SDMT points, 0.5 PASAT points, Table 5).

Baseline Factors Associated With Change in Cognitive Scores

For the SDMT, age, treatment arm, and SDMT score at baseline were significantly associated with the change in scores between baseline and 96 weeks in the linear regression model, whereas sex and disease duration showed no significant association after adjustment for the other covariates. The regression coefficients were −0.17 per year increase in age ($p < 0.0001$), −0.38 per point increase in SDMT at baseline ($p < 0.0001$), and 1.67 for participants in the treatment arm ($p = 0.01$, adjusted R^2 = 0.22). For the PASAT, only age and PASAT score at baseline were significantly associated with the change in scores between baseline and 96 weeks, whereas sex, treatment arm, and disease duration showed no significant associations. The regression coefficients were −0.08 per year increase in age ($p = 0.001$) and −0.28 per point increase in the PASAT at baseline ($p < 0.0001$, adjusted R^2 = 0.17).

Estimate of Regression to the Mean

Studies of cognitive outcome often select for participants with low cognitive performance. In DECIDE, 182 participants had a baseline SDMT Z-score, and 242 participants had a baseline PASAT Z-score of −1.282 or below. Our calculation showed an estimated regression to the mean of 1.52 SDMT and 0.41 PASAT points between the first and second measurements in the selected cases, which suggests a moderate effect of regression to the mean on the results.

Discussion

In this investigation of people with active RRMS, similar to our previous study in SPMS,[11] the cognitive outcomes SDMT

and PASAT show a steady improvement over time. Because the clinical impression is that cognitive function in RRMS steadily declines, it would be reasonable to assume that a useful measure of cognitive function in RRMS would show worsening over time at the group level, even if only a portion of the trial cohort experienced cognitive decline. Contrary to this expectation, our investigation of the DECIDE data set shows that SDMT and PASAT scores steadily increased throughout the 3 years of follow-up, probably due to a practice effect. We cannot exclude that the increase in SDMT and PASAT scores is due to a true treatment effect of both interferon and daclizumab; however, we deem this possibility unlikely because the increase in SDMT and PASAT scores is roughly equal in both trial arms, although daclizumab is the more efficacious treatment. In our previous investigation on cognitive outcome scores in ASCEND, a large placebo-controlled trial in SPMS,[11] the increases over time were also seen in the placebo arm; thus while not ruling out a treatment effect, it adds to the minimization of such an effect. Further research into SDMT and PASAT scores in the placebo arms from placebo-controlled clinical trials and in real-world patient cohorts is necessary to clarify this point.

Another possible explanation of our findings would be regression to the mean, in the sense that the baseline measurements of the cognitive outcomes were unusually low and that the subsequent regression to the mean at later measurements explains the increase in scores over the duration of the trial. Our exploration of this question suggests a moderate amount of regression to the mean in DECIDE. This effect does not appear strong enough to explain the increase in scores on its own. The observed increase in cognitive scores is most likely a combination of a practice effect, regression to the mean, and possibly also a true treatment effect, with the practice effect as the largest contributor.

Table 4 Percentages of Participants With Worsening and Improvement of Cognitive Outcomes Throughout Follow-up

	Week of follow-up					
	24	48	72	96	120	144
SDMT any worsening U	36.9	33.9	29.8	28.6	28.7	27.8
SDMT any worsening 6 MC	21.6	18.6	18.0	18.4	19.7	—
SDMT any improvement U	56.6	61.0	64.8	67.0	67.8	68.9
SDMT any improvement 6 MC	43.8	50.1	54.8	56.8	56.0	—
SDMT 4-pt worsening U	22.4	22.7	19.1	19.1	18.7	20.3
SDMT 4-pt worsening 6 MC	11.9	11.2	11.2	11.3	12.0	—
SDMT 4-pt improvement U	37.2	44.5	49.9	53.6	53.7	55.7
SDMT 4-pt improvement 6 MC	27.4	34.5	39.9	41.5	44.5	—
SDMT 8-pt worsening U	12.8	12.4	11.3	11.8	11.7	10.5
SDMT 8-pt worsening 6 MC	5.8	6.3	6.1	6.8	5.7	—
SDMT 8-pt improvement U	20.1	27.1	31.4	33.5	36.6	37.0
SDMT 8-pt improvement 6 MC	14.2	19.2	21.8	24.3	28.1	—
SDMT 20% worsening U	9.6	9.3	8.7	8.9	7.9	7.8
SDMT 20% worsening 6 MC	3.5	4.7	4.2	4.7	3.6	—
SDMT 20% improvement U	18.8	23.5	26.0	28.9	31.9	31.6
SDMT 20% improvement 6 MC	13.5	17.2	19.4	23.1	24.6	—
PASAT any worsening U	36.6	31.8	29.5	29.3	26.0	25.7
PASAT any worsening 6 MC	19.9	19.4	19.7	18.0	16.1	—
PASAT any improvement U	52.6	57.7	59.4	60.5	62.7	62.3
PASAT any improvement 6 MC	41.4	47.8	50.3	53.6	53.6	—
PASAT 4-pt worsening U	16.7	14.7	12.8	14.0	13.3	13.2
PASAT 4-pt worsening 6 MC	6.5	6.7	7.1	7.5	7.6	—
PASAT 4-pt improvement U	29.0	34.5	37.9	40.8	41.3	42.6
PASAT 4-pt improvement 6 MC	22.4	28.0	31.3	34.6	34.9	—
PASAT 20% worsening U	4.2	4.4	4.0	5.3	4.1	4.9
PASAT 20% worsening 6 MC	1.7	2.0	2.4	2.4	2.3	—
PASAT 20% improvement U	11.2	16.0	16.5	17.6	19.2	22.3
PASAT 20% improvement 6 MC	8.9	12.1	13.0	14.9	17.2	—
T25FW 20% worsening U	10.1	11.1	13.7	16.4	20.0	18.6
T25FW 20% worsening 6 MC	4.5	6.1	9.8	12.4	13.7	—
T25FW 20% improvement U	6.4	8.6	9.0	10.2	11.7	12.9
T25FW 20% improvement 6 MC	3.8	5.4	6.1	8.0	8.5	—

Abbreviations: PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test; T25FW = timed 25-foot walk; U = unconfirmed; 6 MC = 6 mo confirmed.

To get a better understanding of this issue, we contrasted change on the SDMT and PASAT by several definitions with change on the T25FW, a responsive and reliable measure of physical disability in RRMS.[15] Our intention here was not to suggest that cognitive performance and physical disability are closely correlated, but to compare the SDMT and PASAT with a reliable outcome measure that consistently behaves in keeping with the clinical impression and patient experience of

Table 5 Mean Raw Scores and Change From Baseline in Cognitive Outcomes by Baseline Cognitive Performance

		Week of follow-up						
		Baseline	24	48	72	96	120	144
Baseline SDMT below average	SDMT	35.7, 9.4	39.6, 12.1	41.3, 12.9	42.8, 13.1	42.8, 12.7	43.4, 13.6	43.7, 13.2
	n	899	800	735	672	626	322	174
Baseline SDMT not below average	SDMT	60.3, 11.2	58.6, 12.2	59.5, 12.1	60.0, 12.0	60.4, 12.3	60.4, 11.9	60.8, 12.4
	n	876	830	769	714	650	361	196
Baseline SDMT below average	ΔSDMT	—	4.0, 10.0	5.6, 10.6	7.1, 11.1	7.3, 10.9	8.9, 12.5	9.1, 11.9
	n	—	800	735	672	626	322	174
Baseline SDMT not below average	ΔSDMT	—	−1.7, 12.5	−0.8, 12.5	−0.2, 13.1	0.4, 12.8	1.0, 11.9	1.1, 11.8
	n	—	830	769	714	650	361	196
Baseline SDMT below average	PASAT	42.8, 12.2	44.0, 11.7	45.2, 11.4	45.8, 11.6	46.0, 11.8	46.1, 11.4	46.3, 12.4
	n	870	800	740	679	628	321	176
Baseline SDMT not below average	PASAT	51.5, 8.4	53.6, 7.7	53.5, 7.2	54.1, 7.0	54.2, 7.1	54.3, 7.2	53.7, 7.8
	n	895	838	774	719	651	363	199
Baseline SDMT below average	ΔPASAT	—	1.1, 6.8	2.1, 7.9	2.7, 8.0	2.7, 8.3	3.0, 8.3	3.8, 8.6
	n	—	799	738	677	626	320	175
Baseline SDMT not below average	ΔPASAT	—	1.2, 5.7	2.1, 6.0	2.6, 6.0	2.8, 6.7	3.3, 7.0	3.0, 7.4
	n	—	834	770	715	649	362	198
Baseline PASAT below average	SDMT	41.7, 15.7	42.3, 14.4	43.6, 14.4	44.8, 14.1	44.6, 13.8	45.8, 14.2	46.6, 14.3
	n	817	783	719	649	600	325	186
Baseline PASAT not below average	SDMT	53.7, 14.1	55.0, 13.2	56.5, 13.6	57.4, 13.5	57.8, 13.5	58.0, 13.7	58.6, 13.8
	n	948	883	818	764	704	369	193
Baseline PASAT below average	ΔSDMT	—	0.8, 11.3	2.1, 11.6	3.1, 12.2	3.4, 11.7	4.7, 12.7	4.5, 12.3
	n	—	747	688	623	576	313	176
Baseline PASAT not below average	ΔSDMT	—	1.3, 11.6	2.7, 12.1	3.7, 12.7	4.3, 12.6	4.8, 12.9	5.4, 12.3
	n	—	874	808	755	695	367	191
Baseline PASAT below average	PASAT	37.5, 9.4	40.7, 10.5	42.3, 10.4	42.8, 10.8	43.0, 11.0	43.3, 10.8	43.7, 11.4
	n	864	794	723	657	600	325	191
Baseline PASAT not below average	PASAT	55.6, 3.2	55.0, 5.0	55.4, 5.1	55.9, 4.8	56.0, 5.1	56.2, 4.8	56.1, 5.6
	n	963	890	829	772	710	374	194
Baseline PASAT below average	ΔPASAT	—	3.1, 7.2	4.5, 8.2	5.1, 8.4	5.2, 9.1	5.6, 9.1	6.2, 9.2
	n	—	794	723	657	600	325	191
Baseline PASAT not below average	ΔPASAT	—	−0.6, 4.7	−0.1, 4.9	0.4, 4.7	0.5, 5.1	0.8, 5.1	0.4, 5.7
	n	—	890	829	772	710	374	194

Abbreviations: PASAT = Paced Auditory Serial Addition Test; SDMT = Symbol Digit Modalities Test.
Values are mean with SDs.
ΔSDMT: mean SDMT change from baseline, ΔPASAT: mean PASAT change from baseline.

disability worsening. As expected, worsening events on the T25FW steadily increase, whereas improvement events change little over the 3 years of follow-up. On the T25FW, worsening events always outnumber improvement events. In contrast to this, both the SDMT and PASAT show a steady increase in improvement events throughout follow-up, no matter what

definition of worsening is used, and improvement events always outnumber worsening events. This suggests that the SDMT and PASAT scores do not accurately reflect the steady cognitive decline that people with RRMS experience.

Practice effects can be defined as improvements in performance on a cognitive outcome due to repeated exposure.[19,20] A practice effect on the SDMT has been reported in healthy adults,[21] in people with mild cognitive impairment,[18] and in patients with stroke.[22] More recently, an investigation of 264 individuals with self-declared MS using a smartphone application also found a substantial practice effect for repeated SDMT testing over a 5-week interval.[23] These findings are important to put some reports from clinical trials into perspective. The long-term extension of the BENEFIT trial of interferon beta-1b treatment in people with clinically isolated syndrome reports a 0.9-PASAT point difference between the immediate and delayed treatment groups after 11 years of follow-up and concludes that early treatment had a long-lasting beneficial effect on cognitive outcomes.[24] An analysis in the EXPAND trial of siponimod treatment for secondary progressive MS showed an increase in SDMT scores from baseline of about 1.5 points at 24 months and a decrease in the placebo group of about 1.5 points and concludes that treatment with siponimod beneficially affects cognitive functioning.[25] We warn against a potential overinterpretation of such results, as these reported differences fall well within the range of change from baseline observed in our investigation.

To explore a possible treatment effect of daclizumab on cognitive outcomes, we investigated SDMT and PASAT scores by treatment arm. Throughout follow-up, there was little difference in cognitive performance scores between treatment arms. For example, at 96 weeks, daclizumab-treated trial participants reached a mean SDMT score of 52.4 (SD 15.1) points compared with 51.0 (SD 15.1) points for interferon beta-1a treated trial participants (Table 3), a mean 1.4-point difference. A previous publication on cognitive outcomes in the DECIDE data set using a different methodology interpreted these differences as evidence of daclizumab efficacy when compared with interferon beta-1a for cognitive processing speed improvement.[26] Given the small differences in cognitive performance observed in this study, we again caution against the overinterpretation of such results.

The 4-point threshold for clinically meaningful change on the SDMT that is now generally used is based on a study in 97 people with MS who underwent SDMT measurements at 2 time points 1 or more years apart.[9] In that study, a 4-point raw score decrease was associated with a loss of employment. More recently, another investigation in 166 patients sought to validate thresholds for reliable change and suggested an 8-point cutoff as more meaningful.[10] Neither of these studies corrected for the practice effect of the SDMT. Our investigation shows that the issue of steady improvement on the SDMT persists regardless of the threshold used. Thus, a 4-point decline may underestimate a change attenuated by a

practice effect and may explain why an 8-point decline arose as an alternative cut point. Given the inherent issues of post-baseline increases in scores, regardless of the reasons, more research into the size of the practice effect of the SDMT and PASAT should be performed before recommending a general threshold for clinically meaningful longitudinal change.

Although practice effects are usually understood as a source of error and imprecision, it has also been proposed that the practice effect itself could be seen as a cognitive achievement and may therefore deliver meaningful information.[19,20] Our exploration of the change in cognitive scores by baseline cognitive performance shows that this interpretation may not be relevant in RRMS. In the DECIDE data set, both participants with below-average and with average-and-above cognitive performance at baseline showed increasing cognitive scores during follow-up. For both outcome scores, there was greater improvement in those with below-average performance at baseline compared with those with average-and-above performance. This may be an artifact, in the sense that those with high baseline SDMT and PASAT scores can improve only marginally on their already high baseline score, whereas those with low baseline scores had more room to improve with practice. Our analysis of baseline factors associated with the change in cognitive outcome scores showed that higher age and higher cognitive performance at baseline were associated with a decrease in cognitive scores during follow-up. Sex or disease duration showed no such associations.

Our study has several limitations: although it includes a large cohort of patients with RRMS, it is worth remembering that the DECIDE participants were selected using strict entry criteria. The change in cognitive function of DECIDE participants may therefore differ from unselected patients with RRMS. DECIDE also, in part due to its design, had substantial numbers of individuals without later follow-up visits. The later follow-up time points contained only a fraction of the baseline trial cohort and few of those recruited earlier into the study, which could affect the measurements and introduce bias.

Cognitive decline in all MS phenotypes remains a challenging and patient-relevant issue and an important treatment goal. Further research into longitudinal change in cognitive outcomes in both clinical trial and real-world cohorts is needed to better characterize the performance of such measures and to test and develop new outcome measures to detect true change in clinical trials.

Study Funding

Disclosure

serving on the scientific advisory board and speaker's bureau of Biogen, Celgene, EMD Serono, Genentech, and Novartis. He has received research support from AbbVie Inc, Alexion, Alkermes, Biogen, Celgene, Sanofi Genzyme, Genentech, Novartis, and TG Therapeutics. B. Uitdehaag received consultancy fees and/or research support from Biogen, Sanofi Genzyme, EMD Serono, Novartis, Roche, Teva, and Immunic Therapeutics. E. Strijbis reports no disclosures. G. Cutter served on Data and Safety Monitoring Boards: Applied Therapeutics, AI Therapeutics, AMO Pharma, Astra-Zeneca, Avexis Pharmaceuticals, Biolinerx, Brainstorm Cell Therapeutics, Bristol Meyers Squibb/Celgene, CSL Behring, Galmed Pharmaceuticals, Green Valley Pharma, Horizon Pharmaceuticals, Immunic, Karuna Therapeutics, Mapi Pharmaceuticals LTD, Merck, Mitsubishi Tanabe Pharma Holdings, Opko Biologics, Prothena Biosciences, Novartis, Regeneron, Sanofi-Aventis, Reata Pharmaceuticals, Teva Pharmaceuticals, NHLBI (Protocol Review Committee), University of Texas Southwestern, University of Pennsylvania, and Visioneering Technologies, Inc. Consulting or Advisory Boards: Alexion, Antisense Therapeutics, Biogen, Clinical Trial Solutions LLC, Entelexo Biotherapeutics, Inc., Genzyme, Genentech, GW Pharmaceuticals, Immunic, Immunosis Pty Ltd, Klein Buendel Incorporated, Merck/Serono, Novartis, Perception Neurosciences, Protalix Biotherapeutics, Regeneron, Roche, and SAB Biotherapeutics. Dr. Cutter is employed by the University of Alabama at Birmingham and President of Pythagoras, Inc., a private consulting company located in Birmingham AL. M.W. Koch received consulting fees and travel support from Biogen, Novartis, Roche, Sanofi Genzyme, and EMD Serono. Go to Neurology.org/N for full disclosures.

Publication History

Appendix Authors

Name	Location	Contribution
Nina Castrogiovanni, BSc	Department of Clinical Neurosciences, University of Calgary, Canada	Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; and analysis or interpretation of data
Jop Mostert, MD, PhD	Department of Neurology, Rijnstate Hospital, Arnhem, The Netherlands	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Pavle Repovic, MD, PhD	Multiple Sclerosis Center, Swedish Neuroscience Institute, Seattle, WA	Drafting/revision of the manuscript for content, including medical writing for content

Appendix *(continued)*

Name	Location	Contribution
James D. Bowen, MD	Multiple Sclerosis Center, Swedish Neuroscience Institute, Seattle, WA	Drafting/revision of the manuscript for content, including medical writing for content
Bernard M.J. Uitdehaag, MD, PhD	Department of Neurology, MS Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, The Netherlands	Drafting/revision of the manuscript for content, including medical writing for content
Eva M.M. Strijbis, MD, PhD	Department of Neurology, MS Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, The Netherlands	Drafting/revision of the manuscript for content, including medical writing for content
Gary R. Cutter, PhD	Department of Biostatistics, University of Alabama at Birmingham	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data
Marcus W. Koch, MD, PhD	Department of Clinical Neurosciences, University of Calgary, Canada; Department of Community Health Sciences, University of Calgary, Canada	Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design; and analysis or interpretation of data

References

1. Glanz BI, Holland CM, Gauthier SA, et al. Cognitive dysfunction in patients with clinically isolated syndromes or newly diagnosed multiple sclerosis. *Mult Scler J*. 2007;13(8):1004-1010. doi: 10.1177/1352458507077943
2. Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology*. 1991;41(5):685-691. doi: 10.1212/wnl.41.5.685
3. HoldwickJr D, Wingenfeld SA. The subjective experience of PASAT testing. Does the PASAT induce negative mood? *Arch Clin Neuropsychol*. 1999;14(3):273-284. doi: 10.1016/s0887-6177(98)00021-3
4. Tombaugh TN. A comprehensive review of the paced auditory serial addition test (PASAT). *Arch Clin Neuropsychol*. 2006;21(1):53-76. doi: 10.1016/j.acn.2005.07.006
5. Smith A. *Symbol Digit Modality Test (SDMT) Manual*: Western Psychological Services; 1982.
6. Benedict RH, DeLuca J, Phillips G, LaRocca N, Hudson LD, Rudick R. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler J*. 2017;23(5):721-733. doi: 10.1177/1352458517690821
7. Kalb R, Beier M, Benedict RH, et al. Recommendations for cognitive screening and management in multiple sclerosis care. *Mult Scler J*. 2018;24(13):1665-1680. doi: 10.1177/1352458518803785
8. Strober L, DeLuca J, Benedict RH, et al. Symbol Digit Modalities Test: a valid clinical trial endpoint for measuring cognition in multiple sclerosis. *Mult Scler J*. 2019;25(13):1781-1790. doi: 10.1177/1352458518808204
9. Morrow SA, Drake A, Zivadinov R, Munschauer F, Weinstock-Guttman B, Benedict RHB. Predicting loss of employment over three years in multiple sclerosis: clinically meaningful cognitive decline. *Clin Neuropsychol*. 2010;24(7):1131-1145. doi: 10.1080/13854046.2010.511272
10. Weinstock Z, Morrow S, Conway D, et al. Interpreting change on the Symbol Digit Modalities Test in people with relapsing multiple sclerosis using the reliable change methodology. *Mult Scler J*. 2021;28(7):1101-1111. doi: 10.1177/13524585211049397
11. Koch MW, Mostert J, Repovic P, Bowen JD, Uitdehaag B, Cutter G. Is the Symbol Digit Modalities Test a useful outcome in secondary progressive multiple sclerosis? *Eur J Neurol*. 2021;28(6):2115-2120. doi: 10.1111/ene.14732
12. Kappos L, Wiendl H, Selmaj K, et al. Daclizumab HYP versus interferon beta-1a in relapsing multiple sclerosis. *N Engl J Med*. 2015;373(15):1418-1428. doi: 10.1056/nejmoa1501481

13. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology*. 1983;33(11):1444-1452. doi: 10.1212/wnl.33.11.1444

14. Motl RW, Cohen JA, Benedict R, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler J*. 2017; 23(5):704-710. doi: 10.1177/1352458517690823

15. Koch MW, Mostert JP, Wolinsky JS, Lublin FD, Uitdehaag B, Cutter GR. Comparison of the EDSS, timed 25-foot walk, and the 9-hole Peg test as clinical trial outcomes in relapsing-remitting multiple sclerosis. *Neurology*. 2021;97(16):e1560-e1570. doi: 10.1212/wnl.0000000000012690

16. Schwid SR, Goodman AD, McDermott MP, Bever CF, Cook SD. Quantitative functional measures in MS: what is a reliable change? *Neurology*. 2002;58(8): 1294-1296. doi: 10.1212/wnl.58.8.1294

17. R Core Team. R. *A Language and Environment for Statistical Computing [online]*: R Foundation for Statistical Computing; 2022. Accessed at April 18, 2023. R-project.org/

18. Mathews M, Abner E, Kryscio R, et al. Diagnostic accuracy and practice effects in the National Alzheimer's Coordinating Center Uniform data set neuropsychological battery. *Alzheimer's Demen*. 2014;10(6):675-683. doi: 10.1016/j.jalz.2013.11.007

19. Duff K, Beglinger LJ, Schultz SK, et al. Practice effects in the prediction of long-term cognitive outcome in three patient samples: a novel prognostic index. *Arch Clin Neuropsychol*. 2007;22(1):15-24. doi: 10.1016/j.acn.2006.08.013

20. Wang G, Kennedy RE, Goldberg TE, Fowler ME, Cutter GR, Schneider LS. Using practice effects for targeted trials or sub-group analysis in Alzheimer's disease: how practice effects predict change over time. *PLoS One*. 2020;15(2):e0228064. doi: 10.1371/journal.pone.0228064

21. Pereira DR, Costa P, Cerqueira JJ. Repeated assessment and practice effects of the written symbol digit modalities test using a short inter-test interval. *Arch Clin Neuropsychol*. 2015;30(5):424-434. doi: 10.1093/arclin/acv028

22. Koh C-L, Lu W-S, Chen H-C, Hsueh I-P, Hsieh J-J, Hsieh C-L. Test-retest reliability and practice effect of the oral-format symbol digit modalities test in patients with stroke. *Arch Clin Neuropsychol*. 2011;26(4):356-363. doi: 10.1093/arclin/acr029

23. Woelfle T, Pless S, Wiencierz A, Kappos L, Naegelin Y, Lorscheider J. Practice effects of mobile tests of cognition, dexterity, and mobility on patients with multiple sclerosis: data analysis of a smartphone-based observational study. *J Med Internet Res*. 2021; 23(11):e30394. doi: 10.2196/30394

24. Kappos L, Edan G, Freedman MS, et al. The 11-year long-term follow-up study from the randomized BENEFIT CIS trial. *Neurology*. 2016;87(10):978-987. doi: 10.1212/wnl.0000000000003078

25. Benedict RHB, Tomic D, Cree BA, et al. Siponimod and cognition in secondary progressive multiple sclerosis: EXPAND secondary analyses. *Neurology*. 2021;96(3): e376-e386. doi: 10.1212/wnl.0000000000011275

26. 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 J*. 2018;24(6):795-804. doi: 10.1177/1352458517707345