Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data

Myla D. Goldman, MD, MSc, Nicholas G. LaRocca, PhD, Richard A. Rudick, MD, Lynn D. Hudson, PhD, Peter S. Chin, MD, Gordon S. Francis, MD, Adam Jacobs, PhD, Raj Kapoor, FRCP, Paul M. Matthews, MD, Ellen M. Mowry, MD, MCR, Laura J. Balcer, MD, Michael Panzara, MD, Glenn Phillips, PhD, Bernard M.J. Uitdehaag, MD, and Jeffrey A. Cohen, MD, on behalf of the Multiple Sclerosis Outcome Assessments Consortium

Neurology® 2019;93:e1921-e1931. doi:10.1212/WNL.0000000000008519

Abstract

Objective

We report analyses of a pooled database by the Multiple Sclerosis Outcome Assessments Consortium to evaluate 4 proposed components of a multidimensional test battery.

Methods

Standardized data on 12,776 participants, comprising demographics, multiple sclerosis disease characteristics, Expanded Disability Status Scale (EDSS) score, performance measures, and Short Form–36 Physical Component Summary (SF-36 PCS), were pooled from control and treatment arms of 14 clinical trials. Analyses of Timed 25-Foot Walk (T25FW), 9-Hole Peg Test (9HPT), Low Contrast Letter Acuity (LCLA), and Symbol Digit Modalities Test (SDMT) included measurement properties; construct, convergent, and known group validity; and longitudinal performance of the measures individually and when combined into a multidimensional test battery relative to the EDSS and SF-36 to determine sensitivity and clinical meaningfulness.

Results

The performance measures had excellent test–retest reliability and showed expected differences between subgroups based on disease duration and EDSS level. Progression rates in detecting time to 3-month confirmed worsening were lower for T25FW and 9HPT compared to EDSS, while progression rates for LCLA and SDMT were similar to EDSS. When the 4 measures were analyzed as a multidimensional measure rather than as individual measures, progression on any one performance measure was more sensitive than the EDSS. Worsening on the performance measures analyzed individually or as a multidimensional test battery was associated with clinically meaningful SF-36 PCS score worsening, supporting clinical meaningfulness of designated performance test score worsening.

Conclusion

These results support the use of the 4 proposed performance measures, individually or combined into a multidimensional test battery as study outcome measures.
Recognition of the limitations of existing measures of multiple sclerosis (MS)–related disability led to development of the Multiple Sclerosis Functional Composite (MSFC) as an alternative clinical outcome measure. The MSFC integrated 3 quantitative performance measures: Timed 25-Foot Walk (T25FW), 9-Hole Peg Test (9HPT), and Paced Auditory Serial Addition Test (PASAT). Component test scores were normalized by conversion to z scores using a reference sample and averaged to create the MSFC score. Subsequent utilization of the MSFC in clinical trials and other studies highlighted several advantages: high reliability; coverage of cognition, which is not adequately captured by other outcome measures; a single score on a continuous scale; ease of use; and capacity to be administered by a trained technician rather than a physician. However, while utilization of z scores offered statistical advantages, it impeded MSFC acceptance and adoption. Both regulatory agencies and clinicians had difficulty interpreting the clinical meaningfulness of the MSFC score.

Subsequent research provided guidance around thresholds of MSFC measure variability and clinically meaningful change. In addition, researchers recommended adding Low Contrast Letter Acuity (LCLA), a test of visual impairment validated in MS, and replacing the PASAT with the Symbol Digit Modalities Test (SDMT). In this article, we report the initial analyses of a pooled database to assess measurement properties; sensitivity; construct, convergent, and known group validity; and clinical meaningfulness of 4 performance tests (T25FW, 9HPT, LCLA, and SDMT) individually and combined into a multidimensional outcome measure of MS-related disability.

Methods

Study design

The Multiple Sclerosis Outcome Assessments Consortium (MSOAC) was established in 2012 to accelerate development and validation of improved clinical outcome measures of MS-related disability. The organization of MSOAC and overall approach to develop and validate clinical outcome measures of MS-related disability have been reported previously. Frequent interactions with the European Medicines Agency and US Food and Drug Administration contributed to MSOAC’s research approach. After executing data use agreements, MSOAC obtained prospectively acquired patient-level data from 16 clinical trials including 14,370 participants and combined these into a single database. Fourteen of these trials, comprising 12,776 patients, were used for the analyses reported here.

To allow aggregation of data from distinct datasets, a common data standard was developed through the Coalition for Accelerating Standards and Therapies, and data from each clinical trial were remapped to the Clinical Data Interchange Standards Consortium data standard to create a single pooled data set. The standardized data comprising the control and treatment arms of clinical trials formed the MSOAC database. Data elements included demographics, MS disease characteristics, treatment, relapse data, Expanded Disability Status Scale (EDSS) score, performance measures, and patient-reported outcomes. The MSOAC database does not include the MRI data.

Outcome measures

These analyses focused on 4 performance measures and their relation to other measures in the database: T25FW (short distance walking speed to measure ambulation), 9HPT (dexterity and upper extremity motor function), LCLA (vision), and SDMT (cognitive processing speed and sustained attention). Data on PASAT are presented elsewhere. The EDSS is an ordinal scale ranging from 0 to 10 based on the severity of findings on the neurologic examination, walking ability, and ability to carry out activities of daily living, with higher scores indicating worse disability. The Short Form–36 (SF-36) is a 36-item questionnaire that includes 8 multi-item health concepts (Physical Functioning, Role-Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role-Emotional, and Mental Health). Scores are a mean of subsetted questions and range from 0 to 100; higher scores indicate better health-related quality of life (HRQoL). The SF-36 has 2 summary scales, the Physical Component Summary (PCS) and the Mental Component Summary (MCS), whose calculation produces a T-score, with a mean score of 50 and SD of 10, representing the reference score for the US general population. The Beck Depression Inventory (BDI) is a 21-item self-report measure of depression with scores ranging from 0 to 62 and higher score indicating more severe depression symptoms.

For these analyses, worsening was defined as follows: T25FW (20% increase), 9HPT (20% increase), LCLA with 2.5% contrast (20% or 7-letter decrease), SDMT (4-point increase).
The worsening of disability was defined by the EDSS: T25FW worsened at 18 months compared to 11.4% on EDSS at 18 months. Using a 20% threshold for T25FW, 6.5% worsened compared to EDSS. When the performance tests worsened over 24 months was analyzed (figure 2). The study populations available for each comparison differed, leading to differing proportions with 3-month confirmed worsening on EDSS. Using a 20% threshold for T25FW, 6.5% worsened compared to 20.2% on EDSS. Using a 20% threshold for 9HPT, 2.9% worsened compared to 20.2% on EDSS. Using 7-point threshold for LCLA, 13.1% worsened compared to 16.1% on EDSS. Using 4-point threshold for SDMT, 15.0% worsened at 18 months compared to 11.4% on EDSS at 18 months and 14.5% at 24 months. Thus, progression rates were lower for T25FW and 9HPT compared to that of EDSS, while progression rates for LCLA and SDMT were similar to or higher than compared to EDSS. When the performance tests were combined into a multidimensional outcome measure, the proportion of participants worsening on any one performance test was substantially greater than the proportion worsening on EDSS. When worsening on 2 performance tests was required, sensitivity to disability progression was somewhat reduced compared to the EDSS. The progression events defined by the performance tests were weakly associated with or independent of those defined by the EDSS: T25FW (Cohen κ = 0.02, 95% confidence interval [CI] −0.00 to 0.03), 9HPT (κ = 0.00, 95% CI −0.01 to 0.01), LCLA (κ = 0.11, 95% CI 0.08 to 0.14), and SDMT (κ = −0.02, 95% CI −0.06 to 0.02).

To investigate construct and convergent validity, correlations between the performance measures and EDSS were analyzed (table 3). The T25FW and 9HPT correlated strongly with one another and demonstrated the strongest correlation to the EDSS relative to other performance measures. LCLA and SDMT were weakly correlated to the other performance measures.
measures and EDSS. Between the two, the SDMT had somewhat stronger correlation to the EDSS. Cross-sectional correlations among outcomes at baseline were notably stronger than the correlations among changes from baseline to endpoint, which had a similar pattern of correlative strength (T25FW > 9HPT > SDMT > LCLA), but wholly weaker in magnitude.

Known group validity was assessed as a function of disease duration and disability level (table 4). At baseline, values for all 4 performance measures were better in participants with MS of shorter duration (<10 years since symptom onset) compared to those with disease of longer duration (≥10 years). Similarly, the results on all 4 performance tests were better in participants with lower EDSS scores (0–3.5) vs those with higher EDSS scores (4.0–10).

To explore clinical meaningfulness, correlations were calculated between performance measures and participant self-reported measures of HRQoL and depression (table 3). At baseline, T25FW, 9HPT, and SDMT correlated moderately with SF-36 PCS and significantly but weakly with MCS and BDI. LCLA correlated weakly with SF-36 PCS and MSC, and BDI. Correlations between change baseline to endpoint in the performance measures and change on SF-36 PCS or MCS, or BDI were generally not significant and weak at best (table 3). Among participants with worsening from baseline to endpoint on the T25FW, 9HPT, or SDMT, the mean SF-36 PCS also worsened (p < 0.001, p < 0.001, and p = 0.0308, respectively) (table 5). Similarly, among participants who showed baseline to endpoint worsening on the T25FW, 9HPT, or SDMT, the proportions of participants with 5-point PCS worsening on SF-36 PCS were greater. Non-significant trends were seen for mean PCS change and the proportion with 5-point PCS change among participants who did or did not experience baseline to endpoint worsening on LCLA. Mean PCS worsened among participants with baseline to endpoint worsening in each of 2 groups: those with worsening on any one measure and those worsening on 2 or more performance measures. The SF-36 PCS was improved or stable, respectively, among participants who did not worsen on 1 or on 2 or more performance measures. The SF-36 PCS was improved or stable, respectively, among participants who did not worsen on 1 or on 2 or more performance measures. Similarly, the proportions of participants with 5-point SF-36 PCS worsening were greater among participants who showed baseline to endpoint worsening on 1 or on 2 or more performance measures, suggesting that worsening on any single performance measure was clinically relevant.

Table 1 Baseline characteristics

| Measure                          | N   | Mean (SD) or n (%) | Median (range) |
|---------------------------------|-----|--------------------|----------------|
| Age, y                          | 12,776 | 39.5 (9.92)       | 40.0 (17–72)   |
| Sex, n (%)                      | 12,776 |                 |                |
| Female                          | 8,799 (68.9) |                  |                |
| Male                            | 3,977 (31.1) |                  |                |
| Disease course                  | 12,776 |                 |                |
| Relapsing-remitting             | 10,789 (84.4) |                |                |
| Secondary progressive           | 1,044 (8.2) |                   |                |
| Primary progressive             | 943 (7.4) |                   |                |
| Disease duration, y             | 6,641 | 6.5 (7.26)        | 4.0 (0–48)     |
| EDSS                            | 12,776 | 2.9 (1.63)        | 2.5 (0–8)      |
| T25FW, s                        | 11,649 | 7.6 (9.84)        | 5.4 (1.0–231.0)* |
| 9HPT, s                         | 11,653 | 24.3 (14.30)      | 21.3 (5.0–331.0)* |
| LCLA 2.5% contrast, number correct | 5,669 | 34.6 (11.65)      | 37.0 (0–60)    |
| SDMT, number correct            | 2,583 | 47.9 (15.90)      | 48.0 (0–110)   |
| PCS                             | 7,766 | 41.5 (9.95)       | 40.9 (10–73)   |
| MCS                             | 7,766 | 47.7 (11.53)      | 49.5 (5–74)    |
| BDI                             | 2,824 | 8.9 (8.62)        | 7.0 (0–53)     |

Abbreviations: 9HPT = 9-Hole Peg Test; BDI = Beck Depression Inventory; EDSS = Expanded Disability Status Scale; LCLA = Low Contrast Letter Acuity with 2.5% contrast; MCS = Short Form–36 Mental Component Summary; PCS = Short Form–36 Physical Component Summary; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.

* The extreme values were verified as being present in the database.
Discussion

We present analyses to characterize the measurement properties; sensitivity; construct, convergent, and known group validity; and clinical meaningfulness of 4 performance measures—T25FW, 9HPT, LCLA, and SDMT—to permit use individually or combined into a multidimensional test battery as a primary or co-primary outcome measure. We have assessed the components of the proposed multidimensional test battery in relation both to the EDSS and self-reported measures of health-related quality of life and depression. These results, based on a database of 14 datasets comprising 12,776 participants, represent the largest pooled analysis of prospectively acquired clinical trial data in MS to date. The demographics of the pooled dataset largely reflect the type of patients historically enrolled in MS clinical trials, for which trials in RRMS have predominated.

The distributions of the T25FW and 9HPT were positively skewed, and both measures demonstrated floor effects. The potential for a high proportion of patients to perform these measures as well as can be performed by a healthy control can result in reduced ability to distinguish gradations of performance at the lower end of the scale (demonstrated by far left peaks in figure 1, A and B). Baseline LCLA and SDMT scores were more normally distributed, without evidence of floor or ceiling effects. The T25FW, 9HPT, and LCLA showed no clearcut evidence of practice effects. As is typical of most cognitive measures, the SDMT exhibited some practice effects, but these appeared not to affect the normality of the

![Figure 1](https://example.com/image1.png)

(A) Timed 25-Foot Walk (T25FW) (seconds). (B) 9-Hole Peg Test (9HPT) (seconds). (C) Low Contrast Letter Acuity (LCLA) with 2.5% contrast (number correct). (D) Symbol Digit Modalities Test (SDMT) (number correct).

Table 2 Practice effects and test–retest reliability of performance measures with tests 2–6 each compared to test 1

| Test   | N   | Test 2 | Test 3 | Test 4 | Test 5 | Test 6 | ICC  |
|--------|-----|--------|--------|--------|--------|--------|------|
| T25FW  | 7,971 | 0.08   | 0.08   | 0.05   | 0.08   | 0.13   | 0.71 |
| 9HPT   | 7,973 | 0.02   | 0.00   | −0.03  | −0.04  | 0.00   | 0.84 |
| LCLA   | 4,611 | −0.02  | −0.03  | −0.01  | 0.00   | 0.00   | 0.88 |
| SDMT   | 2,094 | 0.03   | 0.10   | 0.15   | 0.28   | 0.37   | 0.85 |

Abbreviations: 9HPT = 9-Hole Peg Test; ICC = intraclass correlation coefficient (a measure of reliability, higher is better, 1 is the maximum possible score); LCLA = Low Contrast Letter Acuity with 2.5% contrast; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.

The values for tests 2–6 are the regression coefficients for the 2nd to 6th test, expressed as an effect size to make them comparable. For example, with T25FW, the 2nd test was on average 0.08 SDs higher than the first test.
SDMT’s frequency distribution. All 4 performance measures demonstrated good test–retest reliability, indicating they yield reproducible scores if there is no change in the participant’s condition. As a result, changes in a score can be assumed to be due to the participant’s condition rather than measurement variability. These results support the advantageous measurement properties of the 4 performance measures.

These results provide a cautionary note regarding the population for which these measures will be most useful. The majority of participants represented in the pooled dataset had RRMS with relatively mild disability. In turn, the T25FW and 9HPT exhibited floor effects, which may explain the decreased sensitivity of 3-month confirmed worsening of T25FW and 9HPT compared to EDSS. More sensitive tests may be needed in studies enrolling participants with mild gait and upper extremity impairments. One might question whether confirmed worsening on EDSS at the low end of the scale in patients with MS with mild impairment represents increasing disability, or simply new signs on the neurologic examination.

Figure 2 Kaplan-Meier graphs of time to 3-month confirmed disability worsening on performance measures compared to Expanded Disability Status Scale

(A) Timed 25-Foot Walk (T25FW). (B) 9-Hole Peg Test (9HPT). (C) Low Contrast Letter Acuity (LCLA) with 2.5% contrast. (D) Symbol Digit Modalities Test (SDMT). (E) Any 1 or any 2 performance measures.
At baseline, the T25FW and 9HPT had stronger correlations with the EDSS and with each other than with the other performance measures. These results support the construct validity of the T25FW and 9HPT, as both are measures of physical functions that overlap with the EDSS in its lower range (EDSS 0–4.0) as seen in this population. In comparison, LCLA and SDMT correlated less strongly with EDSS and the other performance measures, supporting their additive value, to assess functions not captured by the other performance measures and EDSS. Compared to correlations at baseline, all the correlations for change from baseline to endpoint were much weaker. Cohen kappa coefficients showed that the confirmed worsening events defined by the 4 performance measures were largely independent of those defined by EDSS. Taken together, these results suggest that the 4 performance measures assess overlapping but somewhat different aspects of disability and disability worsening than does the EDSS.

All 4 performance measures were worse in participants with longer MS disease duration and with worse disability measured by EDSS, supporting known group validity. Exploratory analyses were undertaken to assess the clinical meaningfulness of worsening on the performance measures using the SF-36.

Table 3 Correlations between outcome measures

|                | 9HPT       | LCLA       | SDMT       | EDSS       | PCS        | MCS        | BDI        |
|----------------|------------|------------|------------|------------|------------|------------|------------|
| **Baseline correlations** |            |            |            |            |            |            |            |
| T25FW          | 0.52 (0.51 to 0.53) | -0.30 (-0.32 to -0.27) | -0.42 (-0.46 to -0.38) | 0.56 (0.55 to 0.58) | -0.40 (-0.42 to -0.38) | -0.13 (-0.16 to -0.11) | 0.22 (0.18 to 0.26) |
| 9HPT           |            | -0.33 (-0.35 to -0.31) | -0.47 (-0.51 to -0.43) | 0.54 (0.53 to 0.56) | -0.33 (-0.36 to -0.31) | -0.14 (-0.16 to -0.11) | 0.20 (0.16 to 0.24) |
| LCLA           | 0.34 (0.30 to 0.39) | -0.29 (-0.31 to -0.27) | 0.12 (0.09 to 0.14) | 0.19 (0.16 to 0.22) | -0.16 (-0.20 to -0.12) |            |            |
| SDMT           | -0.34 (-0.38 to -0.29) | 0.36 (0.32 to 0.41) | 0.21 (0.16 to 0.26) | -0.20 (-0.24 to -0.15) |            |            |            |

|                | 9HPT change | LCLA change | SDMT change | EDSS change | PCS change | MCS change | BDI change |
|----------------|-------------|-------------|-------------|-------------|------------|------------|------------|
| T25FW change   | 0.30 (0.28 to 0.32) | -0.08 (-0.11 to -0.06) | -0.14 (-0.19 to -0.09) | 0.29 (0.27 to 0.31) | -0.20 (-0.23 to -0.18) | -0.09 (-0.12 to -0.06) | 0.10 (0.05 to 0.14) |
| 9HPT change    | -0.06 (-0.09 to -0.04) | -0.20 (-0.25 to -0.15) | 0.23 (0.22 to 0.25) | -0.16 (-0.19 to -0.13) | -0.07 (-0.10 to -0.05) | 0.11 (0.07 to 0.16) |            |
| LCLA change    | 0.06 (0.01 to 0.11) | -0.11 (-0.13 to -0.08) | 0.02 (-0.01 to 0.05) | 0.06 (0.03 to 0.10) | -0.02 (-0.07 to 0.03) |            |            |
| SDMT change    | -0.12 (-0.16 to -0.08) | 0.00 (-0.01 to 0.05) | 0.06 (0.03 to 0.10) | -0.09 (-0.13 to -0.04) |            |            |            |

Abbreviations: 9HPT = 9-Hole Peg Test; CI = confidence interval; LCLA = Low Contrast letter Acuity with 2.5% contrast; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk. Values are Spearman correlation coefficients (95% CI).

Table 4 Known group analysis of baseline values based on disease duration and disability level

| Disease duration, y | T25FW, s | 9HPT, s | LCLA, number correct | SDMT, number correct | Difference (95% CI), p value | EDSS | 4.0–10 | Difference (95% CI), p Value |
|---------------------|----------|---------|----------------------|----------------------|---------------------------|------|--------|---------------------------|
| <10                 | 7.7      | 24.3    | 33.2                 | 48.5                 | -10,09 to -7.12, p < 0.0001 | 6.1  | 31.8  | N = 11,954, 10.10 (9.48 to 10.72), p < 0.0001 |
| ≥10                 | 13.3     | 29.9    | 30.7                 | 45.2                 | -10.09 to -7.12, p < 0.0001 | 12.7 | 31.8  | N = 11,954, 10.10 (9.48 to 10.72), p < 0.0001 |

Abbreviations: 9HPT = 9-Hole Peg Test; CI = confidence interval; EDSS = Expanded Disability Status Scale; LCLA = Low Contrast Letter Acuity with 2.5% contrast; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.
Table 5 Change in Short Form–36 Physical Component Summary (PCS) in participants with and without worsening on Expanded Disability Status Scale (EDSS) and performance measures

| Disability measure | N      | Absolute change in PCS (SD) among participants with disability measure worsening | Absolute change in PCS (SD) among participants without disability measure worsening | p Value | Percent (95% CI) with 5-point PCS worsening among participants with disability measure worsening | Percent (95% CI) with 5-point PCS worsening among participants without disability measure worsening | Odds ratio (95% CI), p Value |
|--------------------|--------|-----------------------------------------------------------------|---------------------------------------------------------------------------------|---------|------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------|
|                    |        |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
| EDSS               | Total: 7,455 | −2.75 (8.21)                                                 | 0.43 (7.54)                                                                       | p < 0.0001 | 36.6 (34.1 to 39.1)                                       | 20.5 (19.5 to 21.6)                                        | 2.23 (1.98 to 2.53), p < 0.0001 |
|                    | Worse: 1,479 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
|                    | Not worse: 5,976 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
| T25FW (20%)        | Total: 7,455 | −2.18 (7.83)                                                 | 0.37 (7.67)                                                                       | p < 0.0001 | 33.4 (31.1 to 35.7)                                       | 20.9 (19.9 to 22.0)                                        | 1.89 (1.68 to 2.13), p < 0.0001 |
|                    | Worse: 1,666 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
|                    | Not worse: 5,789 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
| 9HPT (20%)         | Total: 7,455 | −2.86 (8.33)                                                 | 0.04 (7.68)                                                                       | p < 0.0001 | 38.6 (34.7 to 42.5)                                       | 22.3 (21.4 to 23.4)                                        | 2.18 (1.84 to 2.59), p < 0.0001 |
|                    | Worse: 622   |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
|                    | Not worse: 6,833 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
| LCLA (7 point)     | Total: 4,678  | 0.03 (7.95)                                                  | 0.38 (7.49)                                                                       | p = 0.2907 | 22.1 (18.8 to 25.7)                                       | 20.0 (18.8 to 21.3)                                        | 1.13 (0.92 to 1.40), p = 0.2662 |
|                    | Worse: 570    |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
|                    | Not worse: 4,108 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
| SDMT (4-point)     | Total: 1,467  | −1.15 (8.19)                                                 | −0.04 (7.69)                                                                       | p = 0.0308 | 28.8 (23.7 to 34.4)                                       | 22.2 (19.9 to 24.7)                                        | 1.42 (1.06 to 1.89), p = 0.0201 |
|                    | Worse: 288    |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
|                    | Not worse: 1,179 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
| Worse on any 1 performance measure (T25FW or 9HPT or LCLA or SDMT) | Total: 7,455 | −1.63 (7.94)                                                 | 0.51 (7.60)                                                                       | p < 0.0001 | 30.8 (29.0 to 32.7)                                       | 20.2 (19.0 to 21.3)                                        | 1.71 (1.58 to 1.97), p < 0.0001 |
|                    | Worse: 2,478  |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |
|                    | Not Worse: 4,977 |                                                                |                                                                                 |         |                                                                                                  |                                                                                                  |                             |

Continued
PCS as an anchor. SF-36 PCS correlated moderately at baseline with T25FW, 9HPT, and SDMT and weakly with LCLA. SF-36 MCS and BDI correlated weakly with all 4 performance measures at baseline. Group aggregate changes from baseline to endpoint in the performance measures and self-report measures correlated weakly or not at all when directionality was not considered. Importantly, for participants experiencing confirmed worsening from baseline to endpoint on the T25FW, 9HPT, and SDMT, the SF-36 PCS was significantly worse, and the likelihood of a 5-point worsening on SF-36 PCS, considered clinically important, was significantly greater. The LCLA results mirrored these findings but with statistically nonsignificant trends. These results indicate that the T25FW, 9HPT, and SDMT assess clinically meaningful aspects of MS-related disability and that the proposed thresholds for clinically meaningful change for each are reasonable. The nonsignificant trend of concomitant worsening in the LCLA and SF-36 PCS provides some support for the clinical meaningfulness of 7-letter change in LCLA. The clinical meaningfulness of the 7-letter LCLA worsening has been demonstrated previously, using validated visual quality of life scales.

Limitations of this work include the availability of somewhat fewer data for LCLA, SDMT, and self-reported measures. Relatively few datasets contained all 4 performance measures and EDSS, limiting the analyses of their relative sensitivity, including in RR vs progressive MS, as a function of baseline disability level, and to capture worsening disability independent of relapse. In addition, our ability to fully explore clinical meaningfulness of the performance measures using self-report measures also was restricted by lack of self-report measures used across studies beyond SF-36 and BDI. Other measures of self-report have been applied to the MS population, but these analyses were limited by the surveys in the existing dataset, that is, the SF-36. Also, although the dataset included the full range of disability, the majority of participants had RRMS and relatively mild disability, with median EDSS of 2.5, reflecting the over-representation of clinical trials in RRMS in the MS field. This point may limit a full understanding of the performance tests in more disabled, progressive populations. Finally, for these analyses, we pooled treatment groups and focused on 3-month confirmed disability worsening (rather than 6-month), which could have affected the results.

These results confirm the advantageous measurement properties of the T25FW, 9HPT, LCLA, and SDMT and support their construct, convergent, and known group validity, and sensitivity, particularly when combined into a multidimensional test battery. The associations with established measures of disability (EDSS) and HRQoL (SF-36) indicate that they evaluate clinically meaningful aspects of MS-related disability. These findings support the use of these measures either alone or together as a multidimensional test battery as primary or key secondary endpoints in MS studies.

Table 5: Change in Short Form–36 Physical Component Summary (PCS) in participants with and without worsening on Expanded Disability Status Scale (EDSS) and performance measures (continued)

| Disability measure | Worse on any 2 or more performance measures | Not Worse: 6,839 |
|--------------------|-------------------------------------------|-----------------|
|                    | Absolute change in PCS (SD) among participants with EDSS worsening | Absolute change in PCS (SD) among participants without EDSS worsening |
|                    | N: 7,455 | n/a | 2.43 (8.26) |
|                    | Percent (95% CI) | 35.4 (31.6 to 39.3) | 32.6 |
|                    | p Value | < 0.0001 | 0.12 |
|                    | Odds ratio (95% CI) | 1.57 (1.27 to 2.33) | 1.87 (1.57 to 2.23) |
|                    | p Value | < 0.0001 | 0.01 |

Abbreviations: 9HPT = 9-Hole Peg Test; CI = confidence interval; LCLA = Low Contrast Letter Acuity with 2.5% contrast; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.
Acknowledgment
P.M.M. acknowledges support for activities related to this work from the NIHR Imperial Biomedical Research Centre.

Study funding
MSOAC is funded through the National Multiple Sclerosis Society grant RG 4869-A-1 to the Critical Path Institute, supplemented by annual dues from industry sponsors.

Disclosure
M. Goldman reports grant support from Biogen, MedDay, and Novartis Pharmaceuticals and consulting fees from Adams, EMD Serono, Novartis Pharmaceuticals, Sanofi Genzyme, and TEVA Pharmaceuticals. N. LaRocca is a salaried employee of the National MS Society. R. Rudick is a salaried employee of Biogen. L. Hudson is a salaried employee of the nonprofit Critical Path Institute. P. Chin is a salaried employee of Genentech. G. Francis reports that he was a salaried employee of Novartis at inception of the project but is now an independent consultant including for Novartis, Genentech/Roche, EMD Serono, and GeNeuro. A. Jacobs is a salaried employee of Premier Research, which was paid a commercial fee by the Critical Path Institute for analysis of the data described in this article. R. Kapoor receives support from the UK National Institute of Health Research UCL/H Biomedical Research Centre and fees for consultancies, lectures, and/or support for travel to medical meetings from Actelion, Biogen, Genzyme, Novartis, Roche, and Teva. P. Matthews has received honoraria from Biogen, Roche, Novartis, Evelo, and Ipsen Pharmaceuticals for advisory board participation, research grants or studentships from Biogen, Nodthera, Nestle, and GE Healthcare, and support for educational meetings from Biogen and Novartis. E. Mowry reports serving as site PI for clinical trials and studies sponsored by Biogen, Roche, Novartis, Evelo, and Ipsen Pharmaceuticals for advisory board participation, and SunPharma. She has research grants or studentships support from Sanofi Genzyme and Biogen, Nodthera, Nestle, and GE Healthcare and support for educational meetings for investigator-initiated trials, and she receives free medication for a clinical trial, of which she is PI, from Teva Neuroscience. She receives royalties for editorial duties from Biogen and Novartis. M. Panzara was an employee of Sanofi Genzyme during the project. G. Phillips was previously and employee of and is currently a stockholder of Biogen. L. Baler reports research funding from Biogen. B. Uitdehaag reports consultancy fees from Sanofi-Genzyme, Biogen Idec, Teva, Merck Serono, and Roche. J. Cohen reports personal compensation for consulting for Alkermes, Biogen, Convelo, EMD Serono, ERT, Gossamer Bio, Mapi, Novartis, and ProValuate; speaking for Mylan and Synthon; and serving as an Editor of Multiple Sclerosis Journal. Go to Neurology.org/N for full disclosures.

Publication history
Received by Neurology February 19, 2019. Accepted in final form June 24, 2019.

Appendix Authors

| Name                  | Location                        | Role          | Contribution                                                                 |
|-----------------------|---------------------------------|---------------|------------------------------------------------------------------------------|
| Myla D. Goldman, MD, MSc | University of Virginia, Charlottesville | Author      | Conceptualized and designed the study, interpreted the results, prepared the first draft, reviewed and edited the manuscript for intellectual content |
| Nicholas G. LaRocca, PhD | National Multiple Sclerosis Society, New York, NY | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Richard A. Rudick, MD | Biogen, Cambridge, MA            | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Lynn D. Hudson, PhD  | Critical Path Institute, Tucson, AZ | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Peter Chin, MD        | Genentech, San Francisco, CA     | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Gordon S. Francis, MD | At onset of MSOAC, employee of Novartis; currently independent neurology clinical development consultant, San Francisco, CA | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Adam Jacobs, PhD      | Premier Research, Quincy, MA     | Author         | Conceptualized and designed the study, performed the statistical analyses, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Raj Kapoor, FRCP      | UCL Institute of Neurology, London, UK | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
| Paul M. Matthews, MD  | Imperial College, London, UK     | Author         | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content |
Appendix (continued)

| Name                  | Location                  | Role                     | Contribution                                                                                                                                 |
|-----------------------|---------------------------|--------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| Ellen M. Mowry, MD, MCR | Johns Hopkins, Baltimore, MD | Author                  | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content                   |
| Michael Panzara, MD   | Wave Life Sciences, Cambridge, MA | Author                  | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content                  |
| Glenn Phillips, PhD    | Biogen, Cambridge, MA     | Author                  | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content                   |
| Laura J. Balcer, MD, MSCE | New York University, NY | Author                  | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content                  |
| Bernard M.J. Uitdehaag, MD | VU University Medical Center, Amsterdam, the Netherlands | Author                  | Conceptualized and designed the study, interpreted the results, reviewed and edited the manuscript for intellectual content                   |
| Jeffrey A. Cohen, MD   | Cleveland Clinic, OH      | Author                  | Conceptualized and designed the study, interpreted the results, prepared the first draft, reviewed and edited the manuscript for intellectual content |

References
1. Whitaker JN, McFarland HF, Rudge P, Reingold SC. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. Mult Scler 1995;1:37–47.
2. Rudick R, Antel J, Confavreux C, et al. Clinical outcomes assessment in multiple sclerosis. Ann Neurol 1996;40:469–479.
3. Rudick R, Antel J, Confavreux C, et al. Recommendations from the National Multiple Sclerosis Society clinical outcomes assessment task force. Ann Neurol 1997;42:379–382.
4. Schwid SR, Goodman AD, McDermott MP, Bever CF, Cook SD. Quantitative functional measures in MS: what is a reliable change? Neurology 2002;58:1294–1296.
5. Kräg T, Van der Linden F, Nielsen J, Uitdehaag B, Polman C. Clinical impact of 20% worsening on Timed 25-Foot Walk and 9-Hole Peg Test in multiple sclerosis. Mult Scler 2006;12:594–598.
6. van Wissen L, Kräg T, Hoogervorst F, Polman C, Uitdehaag B. Outcome measurement in multiple sclerosis: detection of clinically relevant improvement. Mult Scler 2010;16:604–610.
7. Bosma L, Kräg T, Brieva L, et al. Progression on the Multiple Sclerosis Functional Composite in multiple sclerosis: what is the optimal cut-off for the three components? Mult Scler 2010;16:862–867.
8. Balcer LJ, Miller DH, Reingold SC, Cohen JA. Vision and vision-related outcome measures in multiple sclerosis. Brain 2015;138:11–27.
9. Balcer LJ, Raynowska J, Nolan R, et al. Validity of low-contrast letter acuity as a visual performance outcome measure for multiple sclerosis. Mult Scler J 2017;23:734–747.
10. Benedict RHB, Deluca J, Phillips G, et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler J 2017;23:721–733.
11. Ontaneda D, LaRocca N, Coetzee T, Rudick RA. Revisiting the multiple sclerosis functional composite: proceedings from the National Multiple Sclerosis Society (NMSS) task force on clinical disability outcomes. Mult Scler J 2012;18:1074–1080.
12. Rudick RA, LaRocca N, Hudson LD, MSOAC. Multiple Sclerosis Outcome Assessments Consortium: genesis and initial project plan. Mult Scler J 2014;20:12–17.
13. LaRocca NG, Hudson LD, Rudick R, et al. The MSOAC approach to developing performance outcomes to measure and monitor multiple sclerosis disability. Mult Scler J 2018;24:1469–1484.
14. Multiple Sclerosis Outcome Assessments Consortium and the CFAST Multiple Sclerosis Development Team. Therapeutic area data standards user guide for multiple sclerosis version 1.0 [online]. Available at: cdisc.org/therapeutic#MS. Accessed February 22, 2019.
15. 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;23:704–710.
16. Pern P, Lammers L, Francis G, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. Mult Scler J 2017;23:711–720.
17. Strober L, Deluca J, Benedict RHB, et al. Symbol Digit Modalities Test: a valid clinical trial endpoint for measuring cognition in multiple sclerosis. Mult Scler J 2018 Epub ahead of print.
18. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). Neurology 1983;33:1444–1452.
19. Ware JE, Sherbourne CD. The MOS 36-Item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. Med Care 1992;30:473–483.
20. Adams HP Jr, Venturi P, Shapira J, et al. Development of a new scale for measuring disability in multiple sclerosis: the McDonald bladder scale (MSB). Mult Scler J 2007;13:327–330.
21. Benedict RHB, DeLuca J, Phillips G, et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. Mult Scler J 2017;23:721–733.
22. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571.
23. Cohen JA, Reingold SC, Polman CH, Wolinsky JS. National Advisory Committee on Clinical Trials in Multiple Sclerosis. Disability outcome measures in multiple sclerosis trials: current status and future prospects. Lancet Neurol 2012;11:467–476.
24. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582–592.
25. Fischer JS, Jak AJ, Knicker JE, Rudick RA, Cutter G. Multiple Sclerosis Functional Composite (MSFC). Administration and Scoring Manual. Unitech: National Multiple Sclerosis Society, 2001.
26. Winer BJ. Statistical Principles in Experimental Design. 2nd ed. New York: McGraw-Hill, 1971.
27. Critical Path Institute. Available at: c-path.org. Accessed February 19, 2019.
Disputes & Debates: Editors’ Choice

Editors’ Note: Association of Prestroke Metformin Use, Stroke Severity, and Thrombolysis Outcome

In “Association of prestroke metformin use, stroke severity, and thrombolysis outcome,” Westphal et al. reviewed data from the European Thrombolysis in Ischemic Stroke Patients collaboration and compared stroke severity (NIH Stroke Scale), 3-month functional outcome (modified Rankin score), and mortality in 757 patients with type 2 diabetes who received metformin before stroke and 1,162 patients with type 2 diabetes who did not receive metformin before stroke. The authors reported that patients on metformin had less severe strokes, better functional outcome, and lower mortality and concluded that these findings suggest metformin has a protective effect in this patient population. Although he found the findings interesting, Dasheiff believed that this conclusion was premature because (1) correlation does not imply causation, (2) the statistical analysis did not fully compensate for the myriad confounders, and (3) there is no mention of other diabetic medications such as insulin use, which, in addition to the HbA1c and blood glucose may reflect disease severity. Westphal et al. agreed that causality cannot be inferred but emphasized the strengths of their study including the large sample size and their use of propensity score matching. The authors suggested their findings may precipitate additional studies on the relationship between metformin and outcome after stroke and increased use of metformin in patients with diabetes and vascular risk factors.

Ariane Lewis, MD, and Steven Galetta, MD

Neurology® 2021;96:500. doi:10.1212/WNL.0000000000011565

Reader Response: Association of Prestroke Metformin Use, Stroke Severity, and Thrombolysis Outcome

Richard Dasheiff (Plano, TX)

When I first read the article by Westphal et al., I was excited about the possible positive effects of metformin on stroke. I note that a one-point difference in NIHSS is not clinically significant, a one-point change in modified Rankin Score is somewhat significant, and 13% vs 18% death rate is more clinically significant. However, further critical analysis challenges this interpretation. First, the authors conclude that “[t]his suggests a protective effect of MET [metformin].” Neurology reviewers and editors should admonish authors about inferring causality from inappropriately designed studies. In addition, as the authors report, “The number of confounders in this study was 19; together with the treatment group (MET+/MET−), 20 independent variables resulted.” This can only partially be compensated by their propensity score matching and multiple imputation. I would point out several other problems. For example, missing values included international normalized ratio (INR) on admission (29.7%). How could this be if everyone got IV thrombolysis? Per the article by Schetz et al. cited in reference 8, “Laboratory measures obtained on hospital admission include glucose (mmol/L), blood cell counts (platelets, hemoglobin, and leucocytes), international normalized ratio and partial thromboplastin time/ Not all centers have to provide data on all variables but have given a commitment to add missing variables retrospectively, if considered relevant to answer a specific research question.”

Author disclosures are available upon request (journal@neurology.org).
So, another question is whether everyone in this study had an INR before IV thrombolysis? Finally, perhaps the most important factor is the severity of the subject’s type 2 diabetes. Although blood glucose and HbA1c are surrogates for the patient’s clinical state, there is no assessment of other diabetes comediations. Patients usually move along a spectrum of diet control, oral hypoglycemics, and insulin. The composition of the MET group could be weighted toward insulin dependence, which would make them, on average, treated longer for their diabetes and having a more severe illness. That alone might account for their worse clinical outcomes, and not be a protective effect of metformin.

1. Westphal LP, Widmer R, Held U, et al; Thrombolysis in Ischemic Stroke Patients (TRISP) Study Group. Association of prestroke metformin use, stroke severity, and thrombolysis outcome. Neurology 2020;95:e362–e373.

Copyright © 2021 American Academy of Neurology

Author Response: Association of Prestroke Metformin Use, Stroke Severity, and Thrombolysis Outcome

Laura P. Westphal (Zurich), Ulrike Held (Zurich), Stefan Engelter (Basel, Switzerland), and Susanne Wegener (Zurich)

Neurology® 2021;96:501. doi:10.1212/WNL.0000000000011563

We thank Dr. Dasheiff for his interest in our article.1 We agree that based on observational registry-based data, it is not straightforward to claim causality and impossible to account for all, including unknown, confounders. However, a thorough analysis based on propensity score–matched (PSM) sets of patients is mimicking a randomized experiment, and because the data set was very large, we were able to account for many known potential confounders simultaneously. It is common in real-life databases that there are some observations missing. In our study, we tackled this with a complex approach, in which we performed imputation and PSM repeatedly in multiple iterative steps. The observation that international normalized ratio (INR) values were incomplete is explained by the fact that standard operating procedures for intravenous thrombolysis (IVT) differed across centers. Some centers do not determine INR before IVT in patients with neither anticoagulation nor liver or coagulation disorder. We agree that duration of diabetes, nonpharmacological and other cotreatments may matter. This limitation was mentioned in the study. By accounting for the age difference between metformin (MET)-/MET+ groups with PSM, we intended to address this aspect. Our data may serve as an argument for patients and GPs to continue MET in patients with diabetes and vascular risk factors and spur planning of further trials that may prove the suggested protective effect of MET.

1. Westphal LP, Widmer R, Held U, et al; Thrombolysis in Ischemic Stroke Patients (TRISP) Study Group. Association of prestroke metformin use, stroke severity, and thrombolysis outcome. Neurology 2020;95:e362–e373.

Copyright © 2021 American Academy of Neurology
Sánchez Fernández I, Abend NS, Agadi S, et al. Gaps and opportunities in refractory status epilepticus research in children: a multi-center approach by the Pediatric Status Epilepticus Research Group (pSERG). Seizure 2014;23:87–97.

Madzar D, Geyer A, Knappe RU, et al. Association of seizure duration and outcome in refractory status epilepticus. J Neurol 2016;263:485–491.

Reader Response: Association of Guideline Publication and Delays to Treatment in Pediatric Status Epilepticus

Ana C. Albuja (Providence, RI), Meriem K. Bensalem-Owen (Lexington, KY), and Mauricio F. Villamar (Providence, RI)

Neurology® 2021;96:502. doi:10.1212/WNL.0000000000011561

Sánchez Fernández et al.1 reported that the publication of evidence on delays in treatment of pediatric refractory convulsive status epilepticus (rSE) was not associated with improvement in time to treatment of rSE (TTTSE). They propose interventions to reduce TTTSE. We would like to share a successful model that we implemented. In 2013, we identified delays in TTTSE in our institution.2 We saw no improvement in TTTSE after a dual intervention consisting of educational programs for healthcare professionals and the development of an electronic SE order set.3 This prompted us to develop an SE alert system that was modeled after our “code stroke” protocol.4 When an inpatient is suspected to have SE, the SE alert is activated by calling an operator who simultaneously pages the neurology, pharmacy, the rapid response team, and the bed manager simultaneously when an inpatient is suspected to be in SE. Amengual-Gual et al. applauded them for implementing this initiative. It is imperative to continue to identify quality improvement measures to identify SE and ensure that AEDs are given in a timely fashion to patients with SE because it becomes more resistant over time, which may lead to increased morbidity and mortality.1,2

Ariane Lewis, MD, and Steven Galetta, MD

Neurology® 2021;96:502. doi:10.1212/WNL.0000000000011562

1. Sánchez Fernández I, Abend NS, Agadi S, et al. Gaps and opportunities in refractory status epilepticus research in children: a multi-center approach by the Pediatric Status Epilepticus Research Group (pSERG). Seizure 2014;23:87–97.
2. Madzar D, Geyer A, Knappe RU, et al. Association of seizure duration and outcome in refractory status epilepticus. J Neurol 2016;263:485–491.
Author Response: Association of Guideline Publication and Delays to Treatment in Pediatric Status Epilepticus

Marta Amengual-Gual (Boston), Cristina Barcia Aguilar (Boston), and Tobias Loddenkemper (Boston)

Neurology® 2021;96:503. doi:10.1212/WNL.0000000000011560

We thank Drs. Albuja, Bensalem-Owen, and Villamar for their comment on our study1 and for their crucial contribution, highlighting the developments toward improving status epilepticus treatment in this pilot study in 19 adult patients during inpatient care at a single center2 and our publication. We are entirely aligned with these essential suggestions and applaud the authors for their crucial contribution. Similarly—and simultaneously to the authors’ accomplishment—pediatric investigators were able to improve inpatient treatment times in children at a single center.3 Most ongoing delays continue to occur in patients with intermittent refractory status epilepticus—related to difficulties in assessment whether convulsive status epilepticus persists—and in the outpatient setting.4 Our group is actively working on measures that can identify ongoing seizures—not only at a single center but also for larger populations in multicenter settings and in the outpatient setting.5 We agree that—among other approaches such as implementing novel technologies and quality improvement techniques—learning from previous stroke treatment efforts will play an important role.6 We are grateful for the chance to add this outlook to our current article and congratulate Drs. Albuja, Bensalem-Owen, and Villamar on this innovative milestone toward our shared goal of implementing improved status epilepticus care for our patients.

1. Sánchez Fernández I, Abend NS, Amengual-Gual M, et al. Association of guideline publication and delays to treatment in pediatric status epilepticus. Neurology 2020;95:e1222–e1235.
2. Villamar MF, Cook AM, Ke C, et al. Status epilepticus alert reduces time to administration of second-line antiseizure medications. Neurol Clin Pract 2018;8:486–491.
3. Adam P, Ostendorf AP, Merison K, Wheeler TA, Patel AD. Decreasing seizure treatment time through quality improvement reduces critical care utilization. Pediatr Neurol 2018;85:58–66.
4. Sánchez Fernández I, Gaita-Lein M, Abend NS, et al. Factors associated with treatment delays in pediatric refractory convulsive status epilepticus. Neurology 2018;90:e1692–e1701.
5. Gaita-Lein M, Sánchez Fernández I, Ularte-Campos A, Loddenkemper T, Ostendorf A. Timing in the treatment of status epilepticus: from basics to the clinic. Seizure 2019;68:22–30.
6. Amengual-Gual M, Ularte-Campos A, Loddenkemper T. Status epilepticus prevention, ambulatory monitoring, early seizure detection and prediction in at-risk patients. Seizure 2019;68:31–37.
In the article “Evaluation of Multiple Sclerosis Disability Outcome Measures Using Pooled Clinical Trial Data” by M. Goldman et al.,¹ the text contained some discrepancies within the analyses of the MSOAC data set included in the paper. The authors found that the discrepancies were present because of a coding error that produced inaccurate results in the Kaplan-Meier results. The authors regret the errors. The text below, provided by the authors, describes the corrected material:

From the Authors:

In our publication in Neurology,¹ we analyzed a pooled dataset comprising 12,776 clinical trial participants that had been assembled by the Multiple Sclerosis Outcome Assessments Consortium (MSOAC) to evaluate 4 performance tests as proposed components of a multidimensional test battery. We reported measurement properties; construct, convergent, and known group validity; and longitudinal performance of the Timed 25-Foot Walk (T25FW), 9-Hole Peg Test (9HPT), Low Contrast Letter Acuity (LCLA), and Symbol Digit Modalities Test (SDMT) individually and when combined into a multidimensional test battery compared to the Expanded Disability Status Scale (EDSS) and Short-Form-36 Physical Component Summary. The placebo arm data in the MSOAC database were made publicly available to support research by investigators in the MS field. Following publication of our paper, a colleague, using the publicly available placebo data from MSOAC, contacted us after finding seemingly different results than reported in our paper. To better understand the potential discrepancy in results, we returned to our analysis. Subsequently, errors in statistical program (Statistical Analysis System, SAS) coding for some of our analyses were discovered. These errors resulted in incorrect progression rates for the T25FW and 9HPT, as well as the composite measures of progression based on any 1 or any 2 measures, which included T25FW and 9HPT. In addition, while the graph of progression by LCLA was correct, in the text, we incorrectly gave results for LCLA progression (and its kappa coefficient for agreement with EDSS) using a 20% threshold, instead of the 7-point threshold stated. Herein, we report corrected results.

Table 1 Sensitivity to Change of Performance Tests and Expand Disability Status Scale

| Performance test | Kaplan-Meier estimates of the percentage (95% CI) of trial participants with 3-mo confirmed disability progression over 24 mo | Corrected | EDSS²,³ |
|------------------|---------------------------------------------------------------|-----------|---------|
| T25FW⁴           | 6.5                                                           | 23.8 (22.8–24.8) | 20.2 (19.2–21.1) |
| 9HPT⁴            | 2.9                                                           | 8.6 (8.0–9.3)     | 20.2 (19.2–21.1) |
| LCLA⁵            | 13.1                                                          | 13.8 (12.7–14.9)  | 16.1 (15.1–17.2) |
| SDMT⁶            | 15.0f                                                         | 15.0f (13.4–16.7) | 14.5 (12.9–16.2) |
| Worsening on ≥1 performance test | Not previously reported | 41.5 (38.9–44.1)   | 11.5 (9.9–13.3) |
| Worsening on ≥2 performance tests | Not previously reported | 12.9 (11.2–14.8)  | 11.5 (9.9–13.3) |

Abbreviations: 9HPT = 9-Hole Peg Test; CI = confidence interval; EDSS = Expanded Disability Status Scale; LCLA = Low-Contrast Letter Acuity; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.

² Four-point threshold.
³ 20% threshold.
⁴ 20% threshold.
⁵ Seven-point threshold: the previous paper had erroneously reported results for a 20% threshold.
⁶ Four-point threshold.
⁷ Baseline score 0: 1.5-point increase, baseline score 1.0–5.5: 1-point increase, baseline score ≥6.0: 0.5-point increase.
⁸ The study populations available for each comparison differed, leading to differing proportions with 3-mo confirmed worsening on EDSS.
⁹ The previously reported result for 3-mo confirmed worsening over 18 mo was correct. For consistency, we report the rate over 24 mo here.

Copyright © 2021 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.
To compare the sensitivity to change of the performance measures with EDSS, time from baseline to 3-month confirmed worsening over 24 months was analyzed (table 1). We previously reported that the progression rates were lower for T25FW and 9HPT compared to EDSS. In the corrected analyses, the progression rate remained somewhat lower for 9HPT compared to EDSS, while progression rates for T25FW, LCLA, and SDMT were similar to or higher than the EDSS (results for LCLA and SDMT were the same as before, as the programming error affected only T25FW and 9HPT). When the performance tests were combined into a multidimensional outcome measure, the proportion of participants with confirmed worsening on any 1 performance test was substantially greater than the proportion with confirmed worsening on EDSS. When confirmed worsening on 2 performance tests was required, sensitivity to disability progression was similar to that of EDSS. Association of the progression events defined by the performance tests was better correlated with those defined by the EDSS in the corrected analysis, though remained modest (table 2).

Based on the previously reported analyses, we concluded that the results supported the use of the T25FW, 9HPT, LCLA, and SDMT as study outcome measures, both individually or combined into a multidimensional test battery. These corrected results demonstrate better sensitivity to change for the T25FW and 9HPT than previously reported, and further support our original conclusion.

**Reference**

1. Goldman MD, LaRocca N, Rudick R, et al. Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data. Neurology 2019;93:e1921–e1931.

### Table 2 Associations of 3-Month Confirmed Progression Events Defined by Performance Tests and by Expanded Disability Status Scale

| Comparison       | Cohen's κ (95% CI) Previously reported | Corrected                   |
|------------------|----------------------------------------|------------------------------|
| T25FW<sup>a</sup> vs EDSS<sup>a</sup> | 0.02 (-0.00 to 0.03)                  | 0.37 (0.35 to 0.39)          |
| 9HPT<sup>b</sup> vs EDSS<sup>b</sup> | 0.00 (-0.01 to 0.01)                  | 0.25 (0.22 to 0.27)          |
| LCLA<sup>c</sup> vs EDSS<sup>e</sup> | 0.11 (0.08 to 0.14)<sup>f</sup>       | 0.10 (0.07 to 0.13)          |
| SDMT<sup>d</sup> vs EDSS<sup>e</sup> | -0.02 (-0.06 to 0.02)                 | -0.02 (-0.06 to 0.02)       |

Abbreviations: 9HPT = 9-Hole Peg Test; CI = confidence interval; EDSS = Expanded Disability Status Scale; LCLA = Low-Contrast Letter Acuity; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.

<sup>a</sup> 20% threshold.
<sup>b</sup> 20% threshold.
<sup>c</sup> Seven-point threshold (results had previously been given for a 20% threshold).
<sup>d</sup> Four-point threshold.
<sup>e</sup> Baseline score 0: 1.5-point increase, baseline score 1.0–5.5: 1-point increase, baseline score ≥6.0: 0.5-point increase.
<sup>f</sup> The previously reported association between progression events defined by 20% worsening on LCLA and EDSS was correct. For consistency, the association between progression events defined by 7-point worsening on LCLA and EDSS is reported here.