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, et al., on behalf of the Multiple Sclerosis Outcome Assessments Consortium

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

Study objective and summary result
This study evaluated a 4-test multidimensional outcome measure for multiple sclerosis (MS)-related disability, and it obtained results supportive of the utility of this measure.

What is known and what this paper adds
Clinical research into MS has been hampered by the inadequacies of various measures of MS-related disability, including the MS Functional Composite (MSFC) tool. This investigation provides evidence for the utility of an alternative multidimensional outcome measure derived from modifications to the MSFC.

Participants and setting
The investigators pooled data from 12,776 patients with MS who participated in the control and treatment arms of 14 clinical trials.

Design, size, and duration
Four performance measures were assessed: the Timed 25-Foot Walk test, the 9-Hole Peg Test, the Low-Contrast Letter Acuity test, and the Symbol Digit Modalities Test. These tests were combined into a multidimensional test battery, and the investigators assessed the utility of this battery as an outcomes measure for MS-related disability. The Expanded Disability Status Scale (EDSS) and the Short Form–36 Physical Component Summary (SF-36 PCS) were the reference tests for analyses of sensitivity and clinical meaningfulness. Test-specific definitions of score-worsening were applied.

Primary outcome measures
The primary outcome was the test battery’s utility as a measure of clinically meaningful worsening of MS-related disability.

Main results and the role of chance
When the performance measures were combined into a multidimensional measure, the observation of progression on any single measure was a more sensitive marker for clinical worsening than was the EDSS, and score-worsening on the multidimensional measure was associated clinically meaningful worsening of SF-36 PCS scores (p < 0.0001).

Bias, confounding, and other reasons for caution
Relatively few datasets contained data for all 4 performance measures of interest.

Generalizability to other populations
The pooling of data from many patients enrolled in various studies favors the generalizability of the results.

Study funding/potential competing interests
This study was funded by the National MS Society. Some authors report receiving funding, various honoraria and fees, travel expenses, studentships, and employment from healthcare companies; being employed by the National MS Society; receiving funding from the UK government; serving as investigators on industry-sponsored clinical trials; owning stock in Biogen; and serving as an editor for the Multiple Sclerosis Journal. Go to Neurology.org/N for full disclosures.

Table Worsening on the multidimensional measure and worsening on the SF-36 PCS

<table>
<thead>
<tr>
<th>Worsening criterion</th>
<th>Mean absolute change in SF-36 PCS scores (SD) in participants:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With worsening</td>
</tr>
<tr>
<td>Worsening on at least 1 performance metric</td>
<td>-1.63 (7.94)</td>
</tr>
<tr>
<td>Worsening on at least 2 performance metrics</td>
<td>-2.43 (8.26)</td>
</tr>
</tbody>
</table>

A draft of the short-form article was written by M. Dalefield, a writer with Editage, a division of Cactus Communications. The authors of the full-length article and the journal editors edited and approved the final version.

Copyright © 2019 American Academy of Neurology
Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data
Neurology 2019;93:e1921-e1931 Published Online before print October 22, 2019
DOI 10.1212/WNL.0000000000008519

This information is current as of October 22, 2019
Readers' 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.00000000000011565

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 Schetiz 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."
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.


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)


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.


Copyright © 2021 American Academy of Neurology


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. 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. 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. This prompted us to develop an SE alert system that was modeled after our “code stroke” protocol. When an inpatient is suspected to have SE, the SE alert is activated by calling an operator who simultaneously pages the neurology and pharmacy house staff, the rapid response team, and the manager responsible for bed assignment. Use of this alert system led to considerable reduction in time to administration of second-line antiseizure medications (22.21 ± 3.44 minutes) compared with standard care (58.30 ± 6.72 minutes; p < 0.0001). Because many hospitals have alert protocols for management of acute stroke, SE alert systems could be replicated with local resources to facilitate timely management and coordinated care of SE in adults and children.
We thank Drs. Albuja, Bensalem-Owen, and Villamar for their comment on our study 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 center 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. 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. 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. 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. 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.

In the article “Evaluation of Multiple Sclerosis Disability Outcome Measures Using Pooled Clinical Trial Data” by M. Goldman et al., 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*, 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 Expanded Disability Status Scale

<table>
<thead>
<tr>
<th>Performance test</th>
<th>Kaplan-Meier estimates of the percentage (95% CI) of trial participants with 3-mo confirmed disability progression over 24 mo</th>
<th>Corrected</th>
<th>EDSS$^e,f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>T25FW$^a$</td>
<td>6.5</td>
<td>23.8 (22.8–24.8)</td>
<td>20.2 (19.2–21.1)</td>
</tr>
<tr>
<td>9HPT$^b$</td>
<td>2.9</td>
<td>8.6 (8.0–9.3)</td>
<td>20.2 (19.2–21.1)</td>
</tr>
<tr>
<td>LCLA$^c$</td>
<td>13.1</td>
<td>13.8 (12.7–14.9)</td>
<td>16.1 (15.1–17.2)</td>
</tr>
<tr>
<td>SDMT$^d$</td>
<td>15.0$^f$</td>
<td>15.0$^f$ (13.4–16.7)</td>
<td>14.5 (12.9–16.2)</td>
</tr>
<tr>
<td>Worsening on ≥1 performance test</td>
<td>Not previously reported</td>
<td>41.5 (38.9–44.1)</td>
<td>11.5 (9.9–13.3)</td>
</tr>
<tr>
<td>Worsening on ≥2 performance tests</td>
<td>Not previously reported</td>
<td>12.9 (11.2–14.8)</td>
<td>11.5 (9.9–13.3)</td>
</tr>
</tbody>
</table>

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.

$^a$ 20% threshold.

$^b$ 20% threshold.

$^c$ Seven-point threshold: the previous paper had erroneously reported results for a 20% threshold.

$^d$ Four-point threshold.

$^e$ Baseline score 0: 1.5-point increase, baseline score 1.0–5.5: 1-point increase, baseline score ≥6.0: 0.5-point increase.

$^f$ The study populations available for each comparison differed, leading to differing proportions with 3-mo confirmed worsening on EDSS.

$^g$ The previously reported result for 3-mo confirmed worsening over 18 mo was correct. For consistency, we report the rate over 24 mo here.
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**

### Table 2

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Previously reported</th>
<th>Corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td>T25FW&lt;sup&gt;a&lt;/sup&gt; vs EDSS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.02 (−0.00 to 0.03)</td>
<td>0.37 (0.35 to 0.39)</td>
</tr>
<tr>
<td>9HPT&lt;sup&gt;b&lt;/sup&gt; vs EDSS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.00 (−0.01 to 0.01)</td>
<td>0.25 (0.22 to 0.27)</td>
</tr>
<tr>
<td>LCLA&lt;sup&gt;c&lt;/sup&gt; vs EDSS&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.11 (0.08 to 0.14)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>0.10 (0.07 to 0.13)</td>
</tr>
<tr>
<td>SDMT&lt;sup&gt;d&lt;/sup&gt; vs EDSS&lt;sup&gt;e&lt;/sup&gt;</td>
<td>−0.02 (−0.06 to 0.02)</td>
<td>−0.02 (−0.06 to 0.02)</td>
</tr>
</tbody>
</table>

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.